charles river analytics

Charles River Analytics Inc.
625 Mount Auburn St.
Cambridge, MA 02138

www.cra.com

p: 617.491.3474
f: 617.868.0780

20 November 2014

Dr. Harold Hawkins
ONR Code 341
Office of Naval Research
875 North Randolph SL
Arlington. VA 22203-1995

Reference: US Navy Contract N00014-12-C-0653: "The Model Analyst's Toolkit: Scientific Model Development, Analysis, and Validation"
Charles River Analytics Contract No. C12186

Subject: Contractor's Quarterly Status Report #9
Reporting Period: 20-August-2014 to 19-November-2014

Dear Dr. Hawkins,

Please find enclosed 1 copy of the Quarterly Status Report for the referenced contract. Please feel free to contact me with any questions regarding this report or the status of the "The Model Analyst's Toolkit: Scientific Model Development, Analysis, and Validation" effort.

Sincerely,

W. Scott Neal Reilly
Principal Investigator

cc: Cheryl Gonzales, DCMA
Annetta Burger, ONR
Whitney McCoy, Charles River Analytics

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
20 NOV 2014		**20-08-2014 to 19-11-2014**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
The Model Analyst's Toolkit:Scientific Model Development, Analysis, and Validation	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Charles River Analytics Inc.,,625 Mount Auburn St.,Cambridge,,MA,02138**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Same as Report (SAR)**	**79**	

Charles River Analytics

Monthly Technical Progress Report No. R12186-09
Reporting Period: August 20, 2014 to November 19, 2014
Government Contract No. N00014-12-C-0653
Charles River Analytics Contract No. C12186

The Model Analyst's Toolkit: Scientific Model Development, Analysis, and Validation

Quarterly Status Report

Principal Investigator: Scott Neal Reilly

Charles River Analytics
625 Mount Auburn Street
Cambridge, MA 02138
617-491-3474

November 20, 2014

1. Executive Summary

The proposed research effort builds on and extends the work of the previous ONR-funded "Validation Coverage Toolkit for HSCB Models" project. The overall objectives of the on-going research program are:

- Help scientists create, analyze, refine, and validate rich scientific models
- Help computational scientists verify the correctness of their implementations of those models
- Help users of scientific models, including decision makers within the US Navy, to use those models correctly and with confidence
- Use a combination of human-driven data visualization and analysis, automated data analysis, and machine learning to leverage human expertise in model building with automated analyses of complex models against large datasets

Specific objectives for the current effort include:

- **Fluid temporal correlation analysis.** Our objective is to design a new method for performing temporally fluid correlation analysis for temporal sets of data and implement the method as a new prototype component within the Model Analyst's Toolkit (MAT) software application.
- **Automated suggestions for model construction and refinement.** Our objective is to design and implement a prototype mechanism that learns from data how factors interact in non-trivial ways in scientific models.
- **Data validation and repair.** Our objective is to design and implement a prototype capability to identify likely errors in data based on anomalies relative to historic data and to use models of historic data to offer suggested repairs.
- **System prototyping.** Our objective is to incorporate all improvements into the MAT software application and make the resulting application available to the government and academic research community for use in scientific modeling projects.
- **Evaluation of applicability to multiple scientific domains.** Our objective is to ensure (and demonstrate) that MAT can be applied to a wide range of scientific domains by identifying and building at least one neurological and/or physiological model and analyze the associated data with MAT, making any extensions to the MAT tool that are needed to support the analysis of such a model.

2. Overview of Problem and Technical Approach

2.1. Summary of the Problem

One of the most powerful things scientists can do is to create models that describe the world around us. Models help scientists organize their theories and suggest additional experiments to run. Validated models also help others in more practical applications. For instance, in the hands of military decision makers, human social cultural behavior (HSCB) models can help predict instability and the socio-political effects of missions, whereas models of the human brain and

mind can help educators and trainers create curricula that more effectively improve the knowledge, skills, and abilities of their pupils.

While there are various software tools that are used by the scientific community to help them develop and analyze their models (e.g., Excel, R, Simulink, Matlab), they are largely so general in purpose (e.g., Excel, R) or so focused on computational models in particular (e.g., Simulink, Matlab), that they are not ideal for rapid model exploration or for use by non-computational scientists. They also largely ignore the problem of validating the models, especially when the models are positing causal claims as most interesting scientific models do. To address this gap, Charles River Analytics undertook the "Validation Coverage Toolkit for HSCB Models" project with ONR. Under this effort, we successfully designed, implemented, informally evaluated, and deployed a tool called the Model Analyst's Toolkit (MAT), which focused on supporting social scientists to visualize and explore data, develop causal models, and validate those models against available data (Neal Reilly, 2010; Neal Reilly, Pfeffer, Barnett et al., 2011, 2010).

As part of the development of the MAT tool, we identified four important extensions to that research program that would further support the scientific modeling process:

- Correlation analyses are still the standard way of identifying relationships between factors in a model, but correlations are fundamentally flawed as a tool for analyzing potentially causal or predictive relationships as they assume instantaneous effects. Even performing correlation analyses with a temporal offsets between streams of data is insufficient as the temporal gap between the causal or predictive event and the following event may not be the same every time (either because of variability in the system being modeled or because of variability introduced by a fixed sampling rate). What we need is a novel way of evaluating the true predictive power across streams of data that can deal with fluid offsets between changes in one stream of data and follow events in the other stream of data.
- Modeling complex phenomena is a fundamentally difficult task. Human intuition and analysis is by far the most effective way of performing this task, but even humans can be overwhelmed by the complexity of modeling the systems they are studying (e.g., socio-political system, human neurophysiology). Automated tools, while not especially good at generating reasonable scientific hypotheses, *are* extremely good at processing large amounts of data. We believe there is an opportunity for computational systems to enhance human scientific inquiry. Under the "Validation Coverage Toolkit for HSCB Models" project, we demonstrated how automated tools could help human scientists to analyze and validate their models against data. We believe a similar approach can be used to help suggest modifications to the human-built models to make them better match the available data. To be useful, however, such automated analyses will need to be rich enough to suggest subtle data interactions that are most likely to be missed by the human scientist. For instance, correlations (especially correlations that take into account fluid temporal displacements) could be used to identify likely relationships between streams of data, but such an approach would miss complex, non-linear relationships between interrelated factors that cannot be effectively analyzed with

simple two-way correlations. For instance, if crime waves are associated with increases in unemployment *or* drops in the police presence, that would be hard to identify with a correlation analysis. We need richer automated data analysis techniques that can extract complex, non-linear, multi-variable relationships between data if we are to effectively suggest model improvements to human scientists.

- Even if a scientific model is sound, if the data sets provided as inputs to the model are unreliable, the results of the model are still suspect. And, unfortunately, data will often be wrong. For instance, HSCB surveys are notoriously unreliable and biased for a variety of reasons, and neurological and physiological data can be corrupted by broken or improperly used sensors. If it were possible to identify when data was unreliable and, ideally, even repair the data, then the models that are using the data could once again be effectively used.
- The MAT tool we developed under the "Validation Coverage Toolkit for HSCB Models" project was focused primarily on assisting social scientists in the analysis, refinement, and validation of HSCB models. In parallel with that effort, however, we also took an opportunity to apply MAT to evaluating neurological and physiological data under the DARPA-funded CRANIUM (Cognitive Readiness Agents for Neural Imaging and Understanding Models) program. We discovered the generality of the MAT tool makes it potentially applicable to a great number of different scientific domains. MAT proved to be a useful, but peripheral tool, in CRANIUM. We believe MAT could be applied to a broader suite of scientific modeling problems than it has been so far.

2.2. Summary of our Approach

To address these identified gaps and opportunities, we are extending MAT's support for model development, analysis, refinement, and validation; enhancing MAT to analyze and repair data; and demonstrating MATs usefulness in additional scientific modeling domains. Our approach encompasses the following four areas, which correspond to the four gaps/opportunities identified in the previous section:

- **Temporally Fluid Correlation Analysis.** We are designing a new method to perform Temporally Fluid Correlational Analysis on temporal sets of data, and we are implementing the method as a new component within the MAT software application. The version of MAT at the beginning of the new effort supported correlation analysis for temporally offset data; it shifts the two data streams being compared by a fixed offset that is based on the sampling rate of the data (i.e., data that is sampled annually will be shifted by one year at a time), performs a standard correlation on the shifted data, plots the correlation value against the amount of the offset, and then repeats the process for the next offset amount. If two data streams are shifted by a fixed offset (e.g., changes in one stream are always followed by a comparable value in the other stream after a fixed time), then this method will find that offset. Under the current effort, we are expanding on this capability to support fluid temporal shifts within the data streams. That is, we are making it possible to identify when the temporal offset between the

change in the first data stream and its effect in the second stream is not a static amount of time.

- **Automated suggestions for model construction and refinement.** We are designing and implementing a mechanism to learn how factors interact in non-trivial ways in scientific models. In particular, we are developing a method for learning disjuncts, conjuncts, and negations. This mechanism starts with the model developed by the scientist user and make recommendations for possible adjustments to make it more complete by performing statistical data mining and machine learning.
- **Data validation and repair.** Recognizing that data contains errors is plausible once we understand the relationships between data sets. That is, if we are able to develop models of the correlations between sets of data, then we can build systems that notice when these correlations do not hold in new data, indicating possible errors in data. For instance, if we know that public sentiment tends to vary similarly between nearby towns, then when one town shows anomalous behavior, we can reasonably suspect problems with the data. There might be local issues that cause the anomaly, but it is, at least, worth noting and bringing to the attention of the user of the data and model. As MAT is designed to help analyze models and recognize inter-data relationships, it is primed to perform exactly this analysis. Existing methods perform similar types of analysis for environmental data (Dereszynski & Dietterich, 2007, 2011). For instance, a broken thermometer can be identified and the data from it even estimated by looking at the temperature readings of nearby thermometers, which will generally be highly correlated.
- **Application to multiple scientific modeling domains.** To ensure (and demonstrate) that MAT can be applied to a wide range of scientific domains, we are identifying and building at least one neurological and/or physiological model and analyzing the associated data with MAT, making any extensions to the MAT tool that are needed to support the analysis of such a model. The initial MAT effort focused on HSCB models; by focusing this effort on harder-science models at much shorter time durations, we believe we can effectively evaluate an interesting range of applications of the MAT tool.

3. Current Activities and Status

During the current reporting period, we focused primarily on two areas. First, we released a new version of the MAT software, so much of the period was devoted to bug fixing, UI improvements, testing, and updating the user manual. Below, we summarize some of the key efforts in that area, and the new user manual is attached as an appendix. Second, we continued to focus on publishing about MAT and getting the word out to the research community. Section 5 describes our efforts in that regard.

Software Improvements for New MAT Release

During the current reporting period, according to our bug tracking system, we fixed 93 issues with the MAT software (some bugs, some user interface improvement, some required functionality for the new public release). Our engineers, scientists, and our in-house QA

engineer and technical writer helped us to find and document these problems, which were fixed by the team. Many of these were related to our transition to a new, more flexible windowing support architecture that provides significant amounts of new flexibility in viewing data and developing models. Figures 1 and 2 provide examples of the old and new systems.

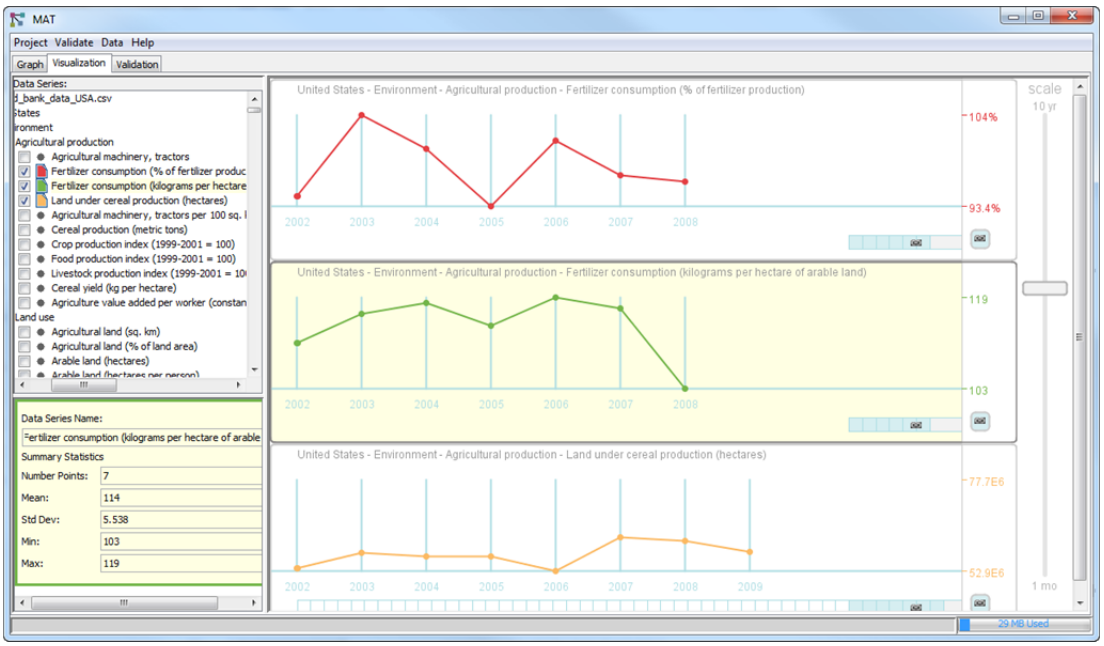


Figure 1. Previous MAT Layout for Data Visualization

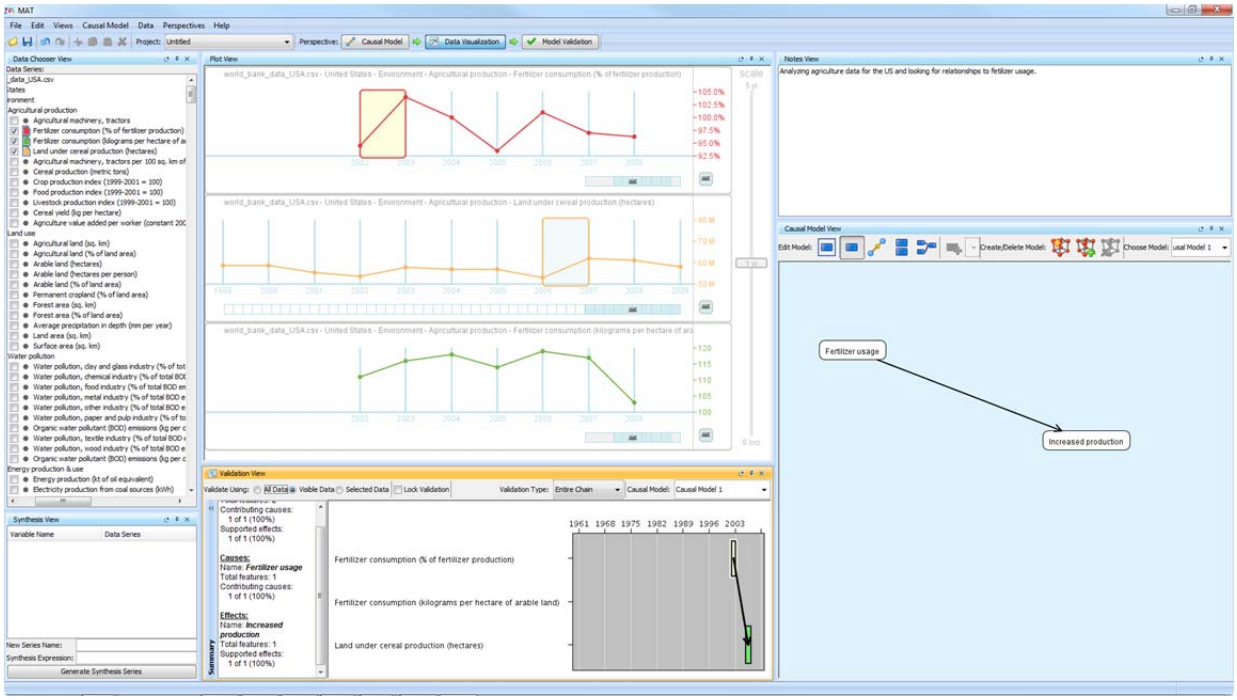


Figure 2. New User-Defined (and Saveable) Layout for Performing Multiple Tasks in Tandem

The new MAT system can be downloaded from our FTP site with a username and password that we provide. We also updated our web site to tell visitors about the new release and to tell them how to request a copy of the new software.

4. Planned Activities

During the upcoming reporting period, we plan to focus on the following tasks:

- Improving the causal-analysis reporting capability of MAT as well as exploring probabilistic-modeling techniques that have been recently developed that might provide additional support for causal analysis.
- Beginning work on data validation.

5. Evaluation and Transition

We continue to focus on making MAT available to the government and academic research communities and to look for opportunities to use MAT on a variety of ongoing research efforts.

During this reporting period, we presented our paper entitled "A Big Data Methodology for Bridging Qualitative and Quantitative Political Science Research" at the American Political Science Association Annual Meeting as part of a panel on Information Technologies in Politics and Political Science. The theme of the 2014 meeting was Politics after the Digital Revolution, examining the way the modern information environment affects not only politics, but the ways in which researchers can study political and social phenomena. In this paper, we present the MAT methodology as a means for both qualitative and quantitative political science researchers to better take advantage of the constantly expanding digital data environment.

The paper was well received and we were informally told that MAT would be nominated for next year's Best Software award. The paper was also on the Social Science Research Network (SSRN) top ten download list for the *Political Methods: Qualitative & Multiple Methods* eJournal.

We also submitted a tutorial proposal to the 2015 International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction (SBP) in March/April that would cover MAT as a tool for mixed-methods scientific discovery. We also submitted a paper abstract on "Tools for Validating Causal and Predictive Claims in Social Science Models" to the 6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015), which was accepted, so we will present that paper in July.

Also, with the new release of MAT, we hope to re-engage a number of researchers who had requested previous versions of the MAT software and to use a press release and social media to alert possible new users to the new software release.

Table 1 summarizes our transition progress to date. We will continue to update this table as we make additional progress and will include it as a regular part of future status reports.

Program	Customer	Comments
On-going efforts		
Tourniquet Master Trainer (TMT) (Phase II SBIR)	US Army's Telemedicine & Advanced Technology Research Center (TATRC)	MAT is being used to visualize and analyze data from sensors on a medical manikin that indicate whether a number of novel medical devices used to combat junctional and inguinal hemorrhaging are being applied properly.
Laparoscopic Surgery Training System (LASTS) (Phase II SBIR)	US Navy's Office of Naval Research (ONR)	Under lasts, Charles River and Caroline Cao at Wright State University are using MAT to analyze data collected from the location of the laproscopic surgery tools tools during an experiment. Surgical tools are instrumented with markers and 3D data is collected on their location as the person performs the task. This is an ongoing Phase II SBIR program.
Cognitive Readiness Agents for Neural Imaging and Understanding Models (CRANIUM) (Phase I SBIR)	US Navy's Office of Naval Research (ONR)	MAT was used to visualize and extract patterns of stress and workload from neuro-physiological data for training systems. This was a Phase I SBIR program that did not progress to Phase II.
Business Intelligence Visualization for Organizational Understanding, Analysis, and Collaboration (BIVOUAC) Phase II SBIR	US Navy's Space and Naval Warfare Systems Command (SPAWAR)	MAT is being evaluated as part of the BIVOUAC SBIR program, which provides data analysis and visualization for Enterprise Resource Planning (ERP) systems for the Navy. This is an ongoing Phase II SBIR program.

Adaptive toolkit for the Assessment and augmentation of Performance by Teams in Real time (ADAPTER) (Phase I SBIR)	US Air Force Research Lab Human Effectiveness Directorate (AFRL/RH)	MAT is being used to analyze neuro-physiological data from cyber operators to evaluate cognitive workload during team-based cyber operations. This is an ongoing Phase II SBIR program.
Anticipated Efforts		
Enhancing Intuitive Decision Making Through Implicit Learning (I2BRC) (ONR Basic Research Challenge BAA)	US Navy's Office of Naval Research (ONR) Charles River is a subcontractor to DSCI MESH Solutions, LLC	The intention is to use MAT to help analyze neuro-physiological data to help better understand how implicit learning and intuitive decision making work. This is an ongoing BAA program. We recently received our first data to review, though the first batch did not include temporal data could leverage MAT.
A system for augmenting training by Monitoring, Extracting, and Decoding Indicators of Cognitive Load (MEDIC)	US Army's Telemedicine & Advanced Technology Research Center (TATRC)	We are evaluating the practicability of using MAT to analyze and visualize neuro-physiological data from combat medic trainees to identify periods of stress and cognitive overload. This is a pending Phase II SBIR program where MAT is being considered for data analysis.
Soldier's Intelligence Fusion Toolkit (SIFT)	US Army Research Laboratory (ARL)	Extend MAT for ARL research objective in high-level information fusion, exploitation, social network analysis and knowledge management research. ARL does not currently have funding for new starts, though we are continuing to engage with them to identify future opportunities for this effort.

Table 1. MAT Transition and Use Progress

In addition we have provided copies of MAT to the following institutions based on their requests for the software: the University of Michigan, Arizona State University, Kansas State University, University of California at Los Angeles, the Naval Medical Research Unit at Wright Patterson Air Force Base, Concordia University (Montreal), the University of Wisconsin, the University of Maryland, and the Air Force Research Laboratory's Human Effectiveness Directorate, the Intelligence Advanced Research Projects Agency (IARPA), and the Joint Advanced Warfighting Division (JAWD).

6. Budget and Project Tracking

As of November 30, 2014, we have spent $706,155, or 76% of our total budget of $928,224, in 71% of the scheduled time. Our current funding is $862,477, so we have spent 82% of our available funding. Note that these numbers include the 26-NOV-2014 funding increment.

Overall, we believe we are in good shape to complete the project on time and on budget.

7. References

Dereszynski, E. W.& Dietterich, T. G. (2011). Spatiotemporal models for data-anomaly detection in dynamic environmental monitoring campaigns. *ACM Transactions on Sensor Networks (TOSN)*, 8.

Dereszynski, E.& Dietterich, T. (2007). Probabilistic models for anomaly detection in remote sensor data streams.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, 424-438.

Granger, C. W. (1980). Testing for causality: a personal viewpoint. *Journal of Economic Dynamics and Control*, 2, 329-352.

Manning, C., Raghavan, R., and Schutze, H. (2008). *Introduction to Information Retrieval*.: Cambridge University Press.

Myers, C.& Rabiner, L. (1981). A level building dynamic time warping algorithm for connected word recognition. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 29, 284-297.

Neal Reilly, S. (2010). *Validation Coverage Toolkit for HSCB Models*. Charles River Analytics Inc.

Neal Reilly, W. S., Pfeffer, A., and Barnett, J. (2010). A Metamodel Description Language for HSCB Modeling. In *Advances in Cross-Cultural Decision Making*.

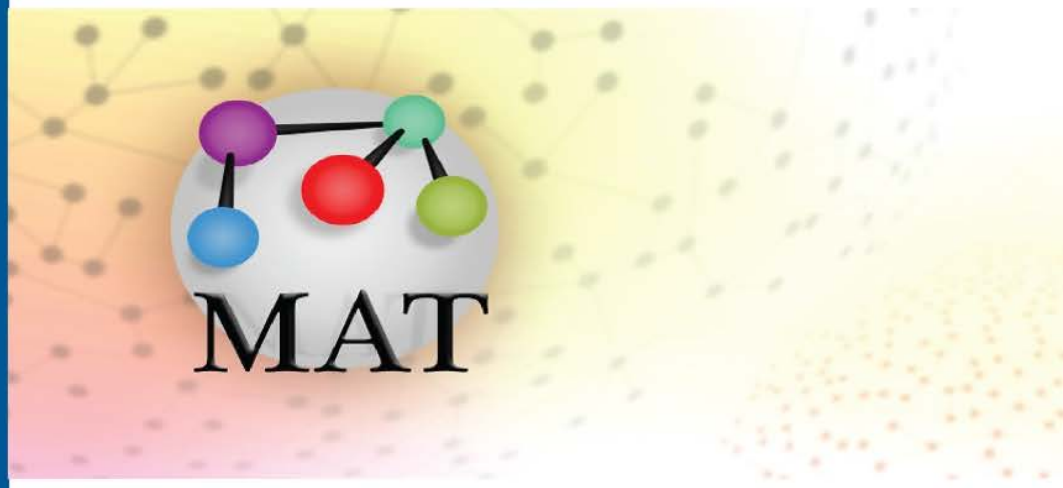
Neal Reilly, W. S., Pfeffer, A., Barnett, J., Chamberlain, J., and Casstevens, R. (2011). A Computational Toolset for Socio-Cultural Data Exploration, Model Refinement, and Model Validation. In *Human Social Culture Behavior (HSCB) Focus*.

Olszewski, R. (2001). *Generalized feature extraction for structural pattern recognition in time-series data*. Carnegie Mellon University, School of Computer Science.

Salvador, S.& Chan, P. (2007). Toward accurate dynamic time warping in linear time and space. *Intelligent Data Analysis*, 11, 561-580.

Sugihara, G., May, R., Ye, H., Hsieh, C., Deyle, E., Fogarty, M., and Munch, S. (2012). Detecting causality in complex ecosystems. *Science*, 338, 496-500.

Model Analyst's Toolkit



User Guide

Version 7.0.0

charles river analytics

Charles River Analytics Inc.
625 Mt. Auburn Street
Cambridge, MA 02138

Examples and Data

All software and related documentation is subject to restrictions on use and disclosure as set forth in the Charles River Analytics Inc. Software License and Services Agreement, with restricted rights for U.S. government users and applicable export regulations.

Companies, names, and data used in examples herein are fictitious unless otherwise noted.

Trademarks and Patents

Charles River Analytics Inc. is registered in the U.S. Patent and Trademark Office. All other trademarks are the property of their respective owners.

NOTICE: This document contains confidential and proprietary information of Charles River Analytics Inc. Use, duplication, or disclosure without the express written consent of Charles River Analytics Inc. is prohibited.

SBIR Data Rights

Charles River Analytics Inc. holds full rights in all technical data and computer software described herein for five years from end of contract, in accordance with DFARS clause 252.227-7018. The Government's rights to use, modify, reproduce, release, perform, display, or disclose technical data or computer software marked with this legend are restricted during the period shown as provided in paragraph (b)(4) of the Rights in Noncommercial Technical Data and Computer Software--Small Business Innovation Research (SBIR) Program clause contained in Contract #W911QY-09-C-0023. No restrictions apply after March 1, 2015. Any reproduction of technical data, computer software, or portions thereof marked with this legend must also reproduce the markings.

Document Information

MAT Model Analyst's Toolkit User Guide, Version 7.0.0
Last updated December 10, 2014

Table of Contents

ABOUT THIS GUIDE

This section provides information about this User's Guide, and the other ways in which Charles River Analytics supports the Model Analyst's Toolkit.

Key topics include:

- Overview
- Typographic conventions
- Feedback

Overview

This guide is designed to allow you to quickly access the information you need to use the Model Analyst's Toolkit. It is also designed to allow you to build your knowledge of the Model Analyst's Toolkit. Use the Table of Contents and Index to quickly find the answer to a specific question, or read the entire book for a complete understanding of all the functionality offered in MAT.

Typographic conventions

Specific conventions are used in this guide to convey additional information about a subject:

Style	Description	Example
`Code`	Code style is used for text that is used literally, appearing exactly as shown. This includes command names, path and file names, and system information.	`E:\MAT\setup.exe`
`Italic code`	Italic code style is used for names of variables that you must provide. For example, you need to supply a value for *`your_file`* in the path name example to the right.	`C:\MAT\data\`*`your_file`*
GUI	GUI style is used to indicate objects in the MAT interface.	the Document field
Bold GUI	Bold GUI style is used to indicate objects with which you interact, such as buttons or menus.	Select **File** > **New** from the menubar. Press the **Enter** key.
Blue GUI	Blue GUI style is used to indicate text you enter into a field.	Enter 10 km in the Range field.

Note Notes highlight information, provide supplementary information, offer time-saving or easier ways to perform the same task, or explain how to prevent errors or data loss. Be sure to read this information carefully.

Feedback about this guide

We appreciate your comments about this guide. Please send your comments, questions, and requests for technical support to mat_project@cra.com.

1 INTRODUCTION

The Model Analyst's Toolkit is a software application that helps researchers and scientists construct and validate models of quantitative theories.

This chapter includes the following key topics:

- About the Model Analyst's Toolkit
- Getting help
- The Model Analyst's Toolkit interface
- About causal models, data visualization, and model validation

About the Model Analyst's Toolkit

The Model Analyst's Toolkit (MAT) is software that supports constructing and validating models of quantitative theories. MAT's features provide researchers with a whole new way to pursue scientific discovery.

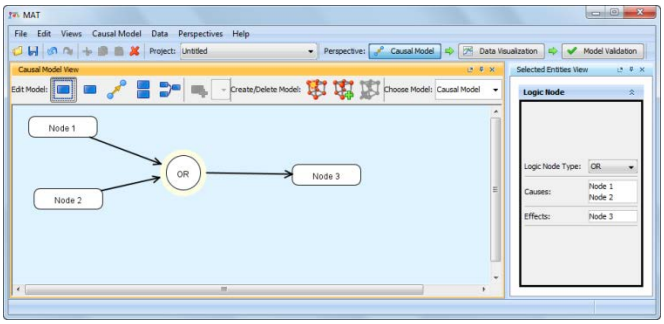
You can use MAT to build a causal model that specifies interconnected causes and effects, and then test that model using any dataset you choose. For example, you might theorize that increased poverty leads to increased crime. MAT lets you combine poverty and crime data to validate your theory.

You can also use MAT to test physical models. For example, you could model the relationships between different points of failure in an engine and then populate the model with engine test data, proving that if a particular component is improved, the engine will be substantially more reliable.

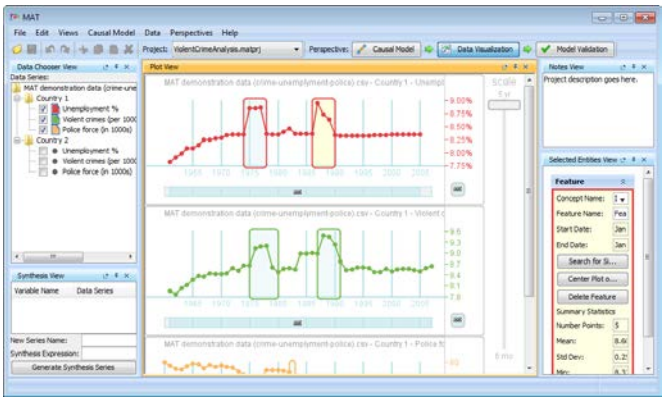
In summary, you can use MAT to turn an idea into a quantitatively provable theory, and visually demonstrate the proof or disproof of that theory. All the interfaces in MAT are easy-to-understand and use a point and click paradigm—no programming or scripting is required.

The process is simple:

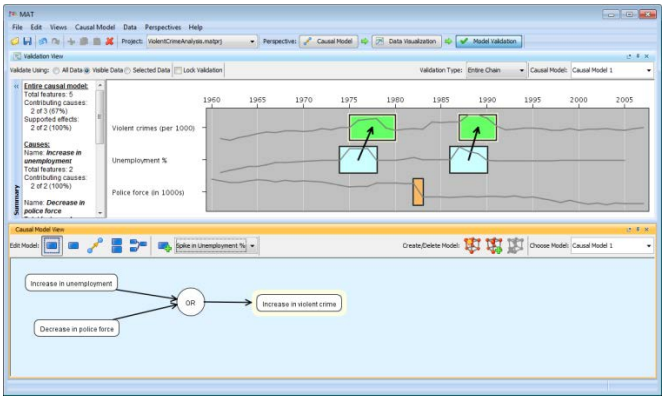
1 Create a graph to represent your model.



2 Associate data features with the concepts in your model.



3 Review MAT's graphical validation results.



Getting help

The Help menu in the MAT menubar offers the following options:

- User Guide

  Select this option to display this user guide.

- About

  Select this option to display MAT support and application information.

Your installation directory includes two sample data files, `example_MAT_data.csv` and `example_MAT_data_withConfig.csv`, located in the `\data` directory. You may find it helpful to open this sample file within the Model Analyst's Toolkit. Many of the screenshots in this User Guide show this sample data.

Please send requests for technical support to mat_project@cra.com.

When contacting us for technical support, the following information may be needed to properly diagnose your issue:

- Content displayed in the About MAT window (displayed when you select Help > About)
- Environment details (operating system)
- `metronome.log` file, located in your MAT 7.0.0 installation folder
- Any log file that begins with a number (for example, `1417697957788.log`) within your MAT 7.0.0 installation/`configuration` folder
- Project file(s) (`.matprj`) open when you encountered the issue
- Brief description of the issue
- Detailed steps to reproduce the issue

Gathering this information before contacting us for support can help us find a resolution more quickly.

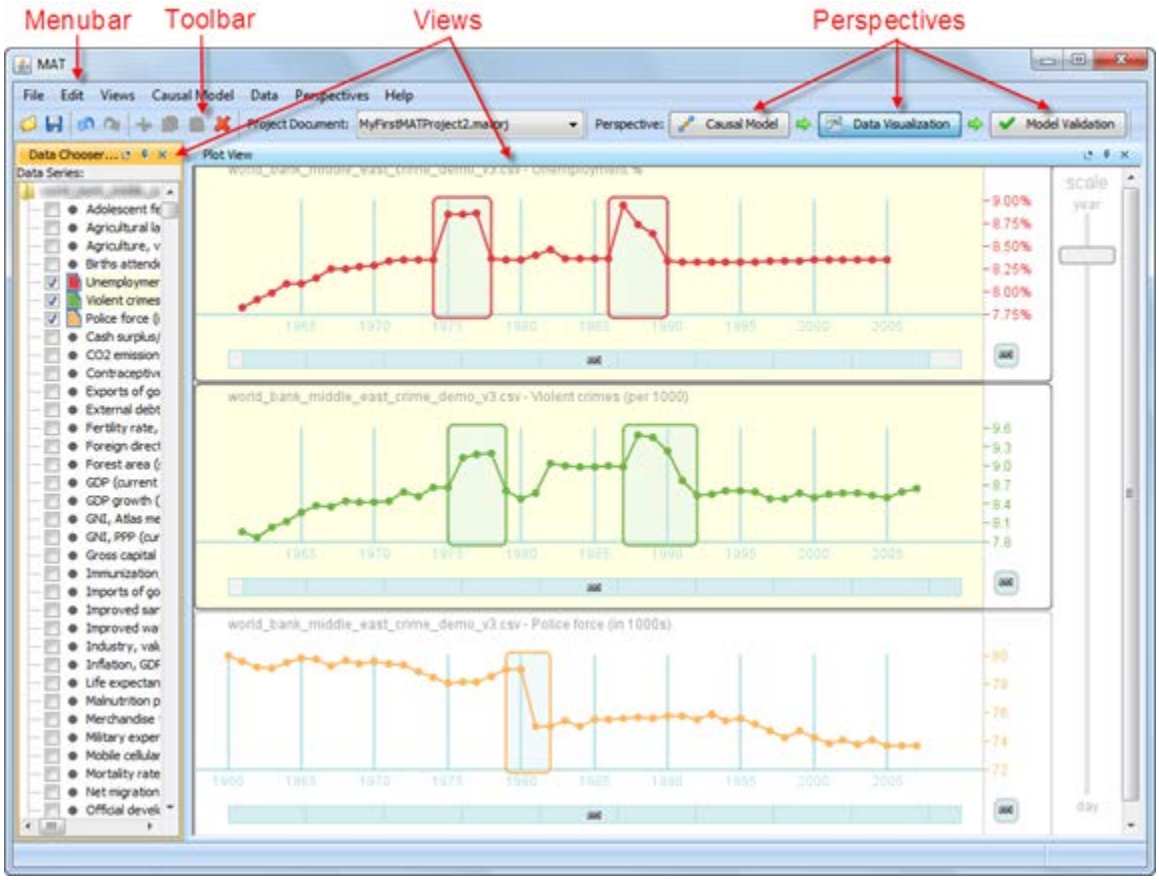
The Model Analyst's Toolkit interface

The MAT graphical user interface (GUI) was designed to reflect Microsoft standards and contains the usual toolbars, buttons, and windows associated with every Microsoft-compliant graphical application.

As a result, only those features of the GUI that relate to performing MAT-specific tasks are explained in this guide. We assume, for example, that you are familiar with standard Windows conventions, such as dragging a window's title bar to move the window, or clicking the close button to close the window.

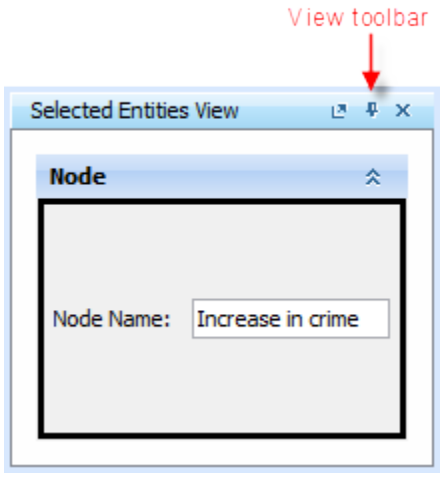
Figure 1 shows the major interface elements in the Model Analyst's Toolkit.

Figure 1 MAT interface elements



Use the standard Windows methods for moving and resizing windows. You can also move and resize views using the toolbar at the top of each view.

Figure 2 Selected Entities view



To rearrange a view

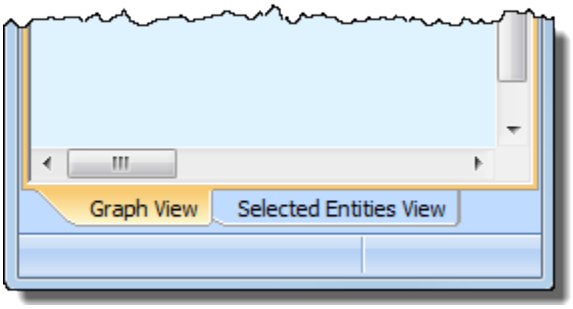
- Float a view by clicking and dragging by the title bar to a new location on your desktop.
- Return a floating view to its previous docked location by clicking .

- Move a floating view to a new docked location by dragging it to an edge of the application window. Possible docking positions are indicated as you pass over them.
- Minimize a docked view to a tab by clicking .
- Expand a minimized view by mousing over or clicking the tab. Return the view to its minimized state by clicking .
- Restore a minimized view by expanding the view and clicking .
- Close a view by clicking .
- If two views share the same docked location, drag a tab to move the view.

To display a view

- Select Views from the menubar and select a view to display.
- Click a tab to switch between views that share the same docked location.

Figure 3 Views sharing a docked location



To return to the default view layout

1. Select Perspectives > Revert to Default Layout from the menubar.
2. Click Yes on the confirmation window.

About causal models, data visualization, and model validation

One of the most powerful things scientists can do is to create models that describe the world around us. Models help scientists organize their theories and suggest additional experiments to run. Validated models also help others in more practical applications. For example, in the hands of military decision makers, human social cultural behavior (HSCB) models can help predict instability and the socio-political effects of missions; models of the human brain and mind can help educators and trainers create curricula that more effectively improve the knowledge, skills, and abilities of their pupils.

Although the scientific community uses various software tools, such as Excel, R, Simulink, and Matlab, to help develop and analyze models, they are largely so general in purpose (Excel, R) or so focused on computational models in particular (Simulink, Matlab), that they are not ideal for rapid model exploration or for use by non-computational scientists. These tools also largely ignore the problem of validating the models, especially when the models are positing causal claims, as most interesting scientific models do. To address this gap, we developed the Model Analyst's Toolkit (MAT), which focuses on enabling social scientists to visualize and explore data, develop causal models, and validate those models against available data.

2 GETTING STARTED

This chapter describes the Model Analyst's Toolkit's technical requirements, and how to install, start, use, exit, and uninstall MAT.

This chapter includes the following topics:

- Technical requirements
- Installing the Model Analyst's Toolkit
- Starting the Model Analyst's Toolkit
- Using the Model Analyst's Toolkit
- Exiting the Model Analyst's Toolkit
- Uninstalling the Model Analyst's Toolkit

Technical requirements

Your system must meet the following requirements to run the Model Analyst's Toolkit:

- Windows XP operating system (or higher) with Service Pack 2 and all critical Windows updates installed
- Java Runtime Environment (JRE) 1.6.0
  This version of the JRE is installed with the Model Analyst's Toolkit if you perform a full installation.
- Pentium IV 3.0 GHz processor or better
  Greater processor speeds will result in faster results when modifying large and complicated networks.
- 1 GB or more of system RAM
- At least 1 GB of free hard drive space
- (Recommended) Screen resolution of 1280x1024 pixels

A network connection is not required for the Model Analyst's Toolkit.

Installing the Model Analyst's Toolkit

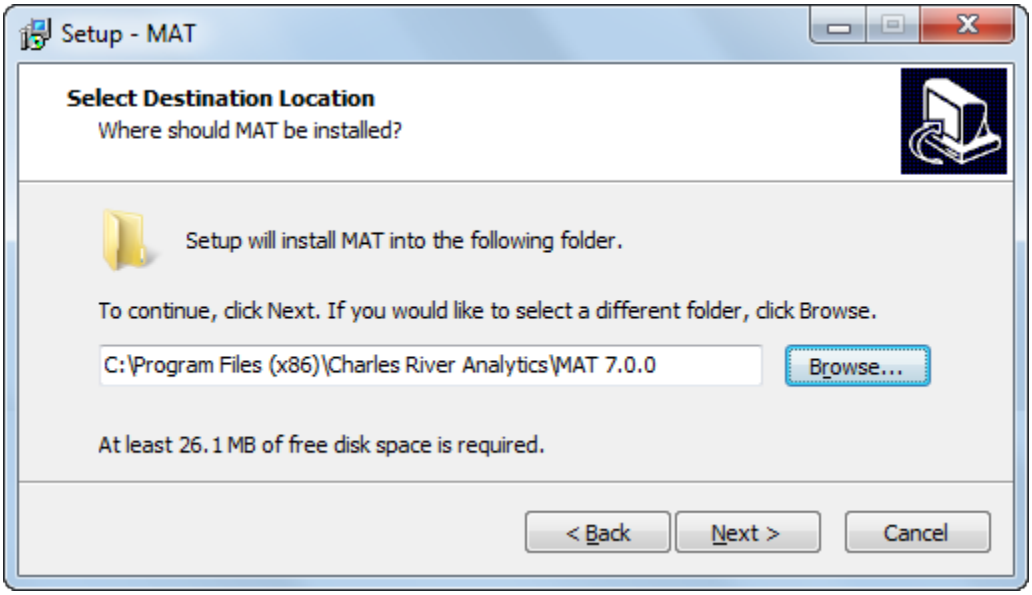
The installation wizard will guide you through each step of Model Analyst's Toolkit installation.

To install the Model Analyst's Toolkit

1 Review the `MAT_License_Agreement.pdf` file within the `licenses` directory in the MAT installation directory.

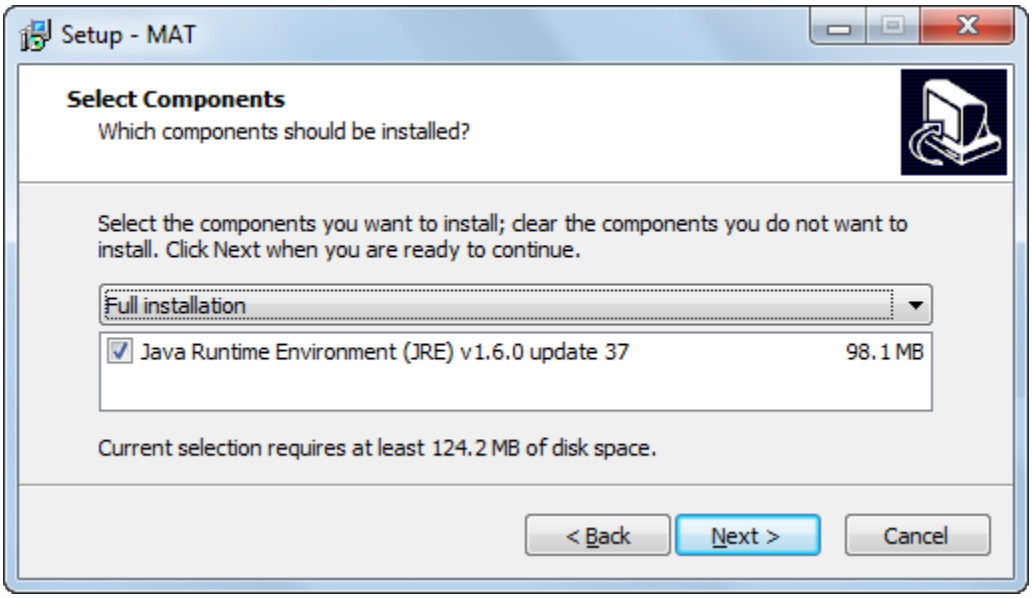
2 Double-click the `MAT 7.0.0 Installer.exe` file to display the MAT setup wizard.

3 Click Next to display the Select Destination Location page.



If you want to select a different folder for the MAT software, click **Browse** to display the Browse for Folder window, navigate to the folder, and click **OK**.

4 Click **Next** to display the Select Components page.

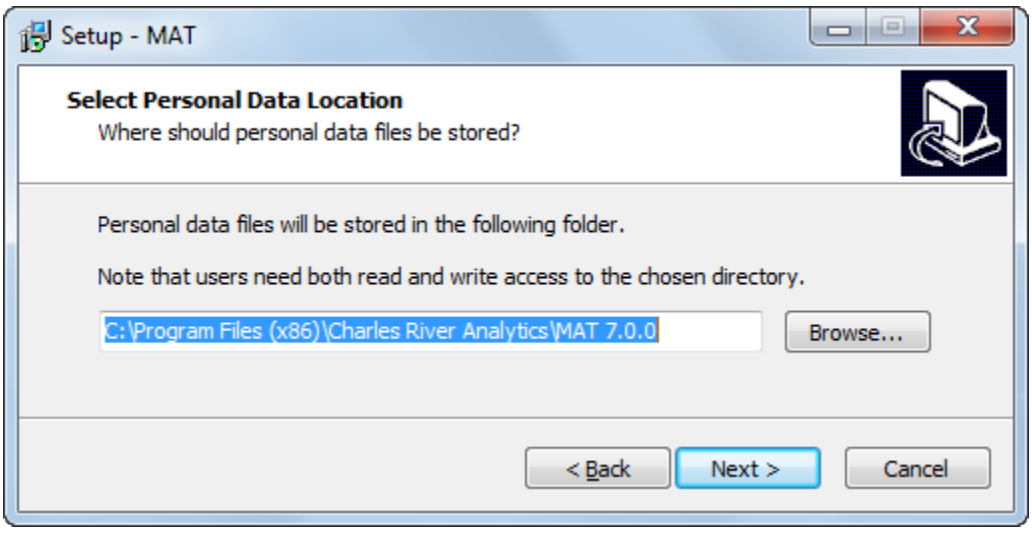


Select the type of installation you want from the drop-down list. Select:

- **Full installation** to include the Java Runtime Environment (JRE)
- **Compact installation** to exclude the JRE

We recommend that you select **Full installation** and leave the **Java Runtime Environment** box checked.

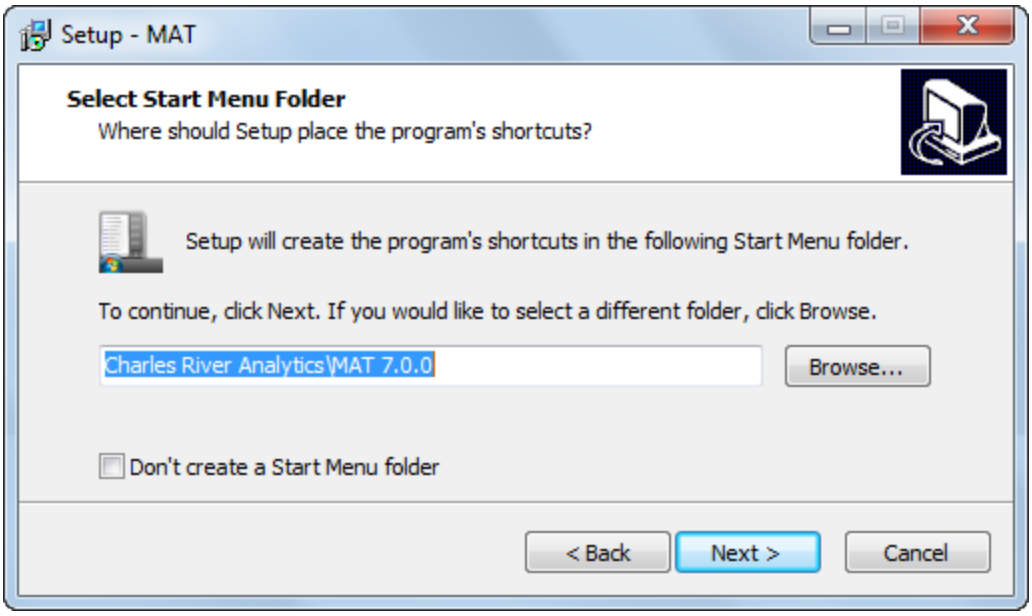
5 Click **Next** to display the Select Personal Data Location page.



MAT stores its configuration files in the selected folder.

If you want to select a different folder for these files, click **Browse** to display the Browse for Folder window, navigate to the folder or create a new folder, and click **OK**.

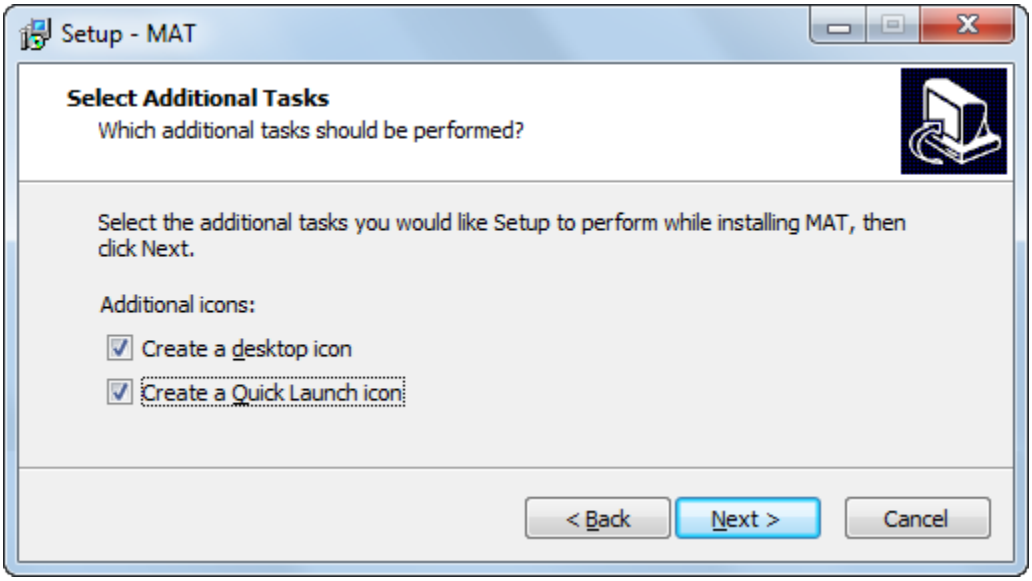
6 Click **Next** to display the Select Start Menu Folder page.



If you want the MAT icon to appear in a different folder in your Windows Start menu , click Browse to display the Browse for Folder window, navigate to the folder in which you want the MAT program icon to appear, and click OK.

If you do not want the MAT application icon to appear in your Windows Start menu, check the Don't create a Start Menu folder box.

7 Click Next to display the Select Additional Tasks page.



Check one or both of the following boxes to create additional MAT application icons:

- Create a desktop icon – Creates a MAT application icon on your desktop
- Create a Quick Launch icon – Creates a MAT application icon on the taskbar for operating systems released before Windows 7, such as Windows NT and Vista.

8 Click Next to display the details of the installation on the Ready to Install page.

9 Click Install to install MAT and display the Completing the MAT Setup page.

A progress bar appears until the installation is complete.

10 Check the Run MAT box and click Finish to close the wizard and run MAT.

Leave the box unchecked to close the wizard without running MAT.

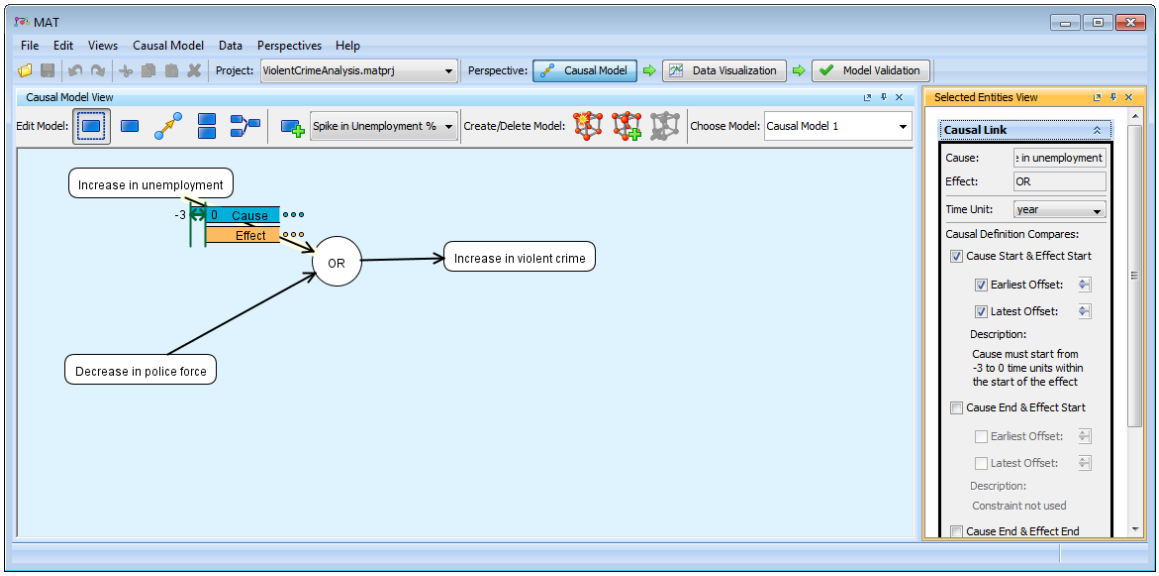
Starting the Model Analyst's Toolkit

You can start the Model Analyst's Toolkit in two ways:

- Select All Programs > Charles River Analytics > MAT 7.0.0 > MAT from the Windows Start menu.
- Double-click `run.bat` in the MAT 7.0.0 folder. This folder's location depends on how you installed MAT.

The first time you open the Model Analyst's Toolkit, the Causal Model perspective is displayed. The next time you start MAT, it displays the perspective you were using when you closed the application.

Figure 4 MAT Causal Model perspective



Using the Model Analyst's Toolkit

You can use the Model Analyst's Toolkit to help you create and validate causal models by following these general steps:

1 Start the Model Analyst's Toolkit.

For more information, see *Starting the Model Analyst's Toolkit*, above.

2 Use a graph to model the concepts from your domain and the causal/predictive relationships between those concepts with a graph.

For more information, see *Working with Causal Models* on page 21.

3 Import, visualize, and explore data relevant to your model.

For more information, see *Importing Data* on page 31 and *Visualizing Data* on page 38.

4 Identify features (events) in the data that relate to the concepts in your model.

For more information, see *Working with data features* on page 48.

5 Analyze your model using the data to determine whether there are problems to be examined.

For more information, see *Model Validation* on page 55.

6 Modify your causal model based on the validation results. You can use MAT's automated data mining techniques to suggest model refinements that will improve the explanatory power of your model.

For more information, see *Generating recommended causal models from the data* on page 23.

Exiting the Model Analyst's Toolkit

You can exit the Model Analyst's Toolkit in several ways:

- Select File > Exit from the MAT menubar.
- Click the close button on the title bar.
- Press Ctrl+Q.

Uninstalling the Model Analyst's Toolkit

Select All Programs > Charles River Analytics > MAT 7.0.0 > Uninstall MAT from the Windows Start menu to uninstall the Model Analyst's Toolkit and all its components.

3 TUTORIAL: WHAT CAUSES INCREASED CRIME?

In this tutorial, you will explore the key features of the Model Analyst's Toolkit using the sample data file provided with MAT. You will create a causal model and use the sample data to explore whether an increase in unemployment and decrease in the police force caused an increase in crime.

This tutorial addresses the following topics:

- Create a causal model
- Import and visualize the data
- Define data features
- Validate the model

Create a causal model

You create your causal model using nodes, logic nodes, and causal links to connect them. In this section of the tutorial you will create three nodes and link them to capture your theory that both an increase in unemployment and a decrease in the police force contribute to an increase in crime.

1 Select All Programs > Charles River Analytics > MAT 7.0.0 > MAT from the Windows Start menu to start MAT and display the Causal Model perspective.

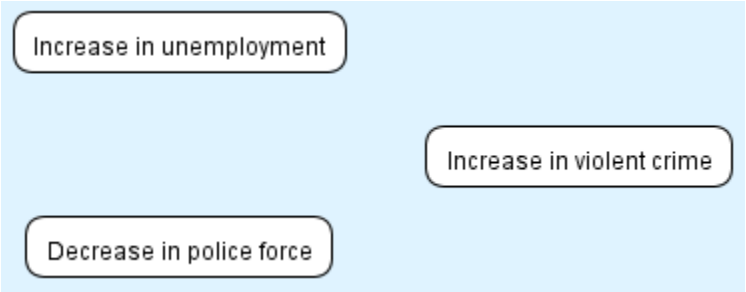
If you do not see this option, double-click `run.bat` in the `MAT 7.0.0` folder. This folder's location depends on where you installed MAT.

If the Causal Model perspective is not displayed, click Causal Model in the MAT toolbar.

2 Click the concept node tool in the Causal Model View toolbar.

3 Click in the Causal Model View window to create three nodes, one for each concept you want to model.

Figure 5 Three concept nodes



4 Click the select tool in the Causal Model View toolbar.

5 Select a node, and enter Increase in unemployment in the Node Name field in the Selected Entities View. Select the other nodes and enter Decrease in police force and Increase in violent crime.

6 Click the causal link tool in the Causal Model View toolbar.

7 Click the Increase in unemployment node, then click the Increase in violent crime node to model their cause and effect relationship.

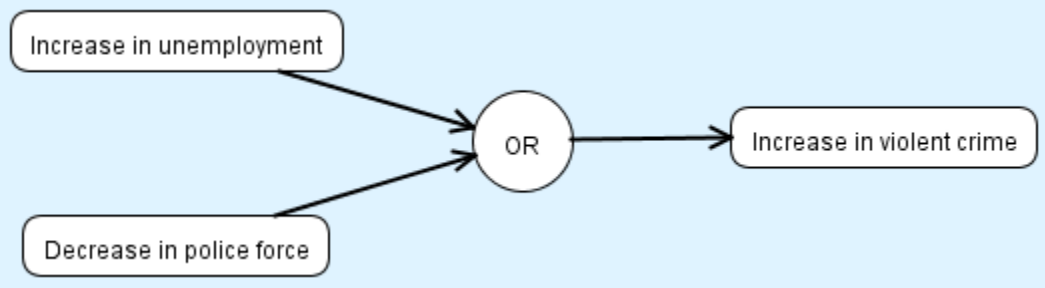
The link is drawn from the first node to the second node and a graphical display of constraints is displayed. The details of the causal link are displayed in the Selected Entities View.

8 Click the Decrease in police force node, then click the Increase in crime node, since your model has two causes for the increase in crime.

MAT displays the Ambiguous Model window. Select OR from the drop-down list and click OK to create a logic node in your causal model.

9 Click the select tool in the Causal Model View toolbar and adjust the nodes so your model looks similar to Figure 6.

Figure 6 Concepts connected with causal links and an OR logic node



For more information about creating and editing causal models, see *Working with Causal Models* on page 21.

Import and visualize the data

In this section of the tutorial, you will import the MAT demonstration data set and visualize the data within the file.

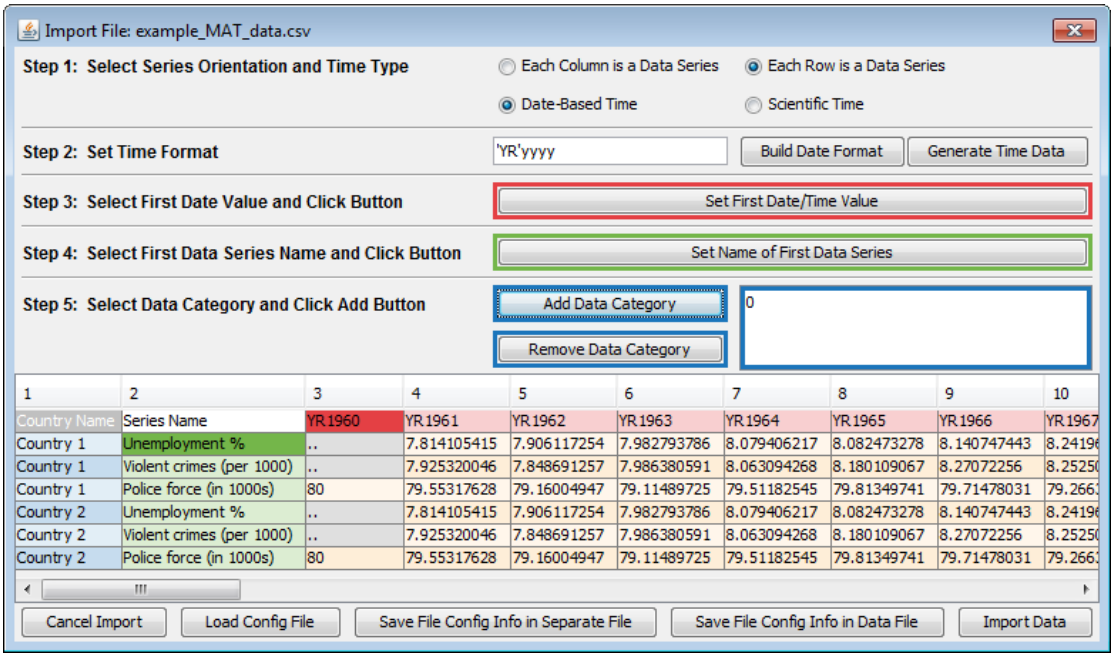
10 Click Data Visualization in the Perspective toolbar to display the Data Visualization perspective.

11 Select File > Import Data from the menubar to display the Open window.

12 Navigate to the `MAT 7.0.0/data` folder in the installation directory. This folder's location depends on where you installed MAT.

13 Double-click the `example_MAT_data.csv` file to display the data in the Import File window.

Figure 7 Example data file configured for import in MAT's import file window



1	2	3	4	5	6	7	8	9	10
Country Name	Series Name	YR1960	YR1961	YR1962	YR1963	YR1964	YR1965	YR1966	YR1967
Country 1	Unemployment %	..	7.814105415	7.906117254	7.982793786	8.079406217	8.082473278	8.140747443	8.2419
Country 1	Violent crimes (per 1000)	..	7.925320046	7.848691257	7.986380591	8.063094268	8.180109067	8.27072256	8.2525
Country 1	Police force (in 1000s)	80	79.55317628	79.16004947	79.11489725	79.51182545	79.81349741	79.71478031	79.266
Country 2	Unemployment %	..	7.814105415	7.906117254	7.982793786	8.079406217	8.082473278	8.140747443	8.2419
Country 2	Violent crimes (per 1000)	..	7.925320046	7.848691257	7.986380591	8.063094268	8.180109067	8.27072256	8.2525
Country 2	Police force (in 1000s)	80	79.55317628	79.16004947	79.11489725	79.51182545	79.81349741	79.71478031	79.266

14 Select Each Row is a Data Series because values for each point in time for the series in this data file appear in a single row.

15 Select Date-Based Time as the time type, since the time values in this data file are based on calendar dates .

16 Enter 'YR'yyyy in the Time Format field to specify that the dates in this data file are given by a four-digit year preceded by the two letters YR.

17 Click the cell in the data table that displays the first time value (YR1960), then click Set First Date/Time Value.

The selected cell is shaded red and the other date/time values in your dataset are shaded pink.

18 Click in the cell in the data table that displays the name of the first data series (Unemployment %), then click Set Name of First Data Series.

The selected cell is shaded green. All data series in your dataset are shaded light green. Date/time values for each data series in the data table are shaded yellow and gray. Gray shading indicates a missing value.

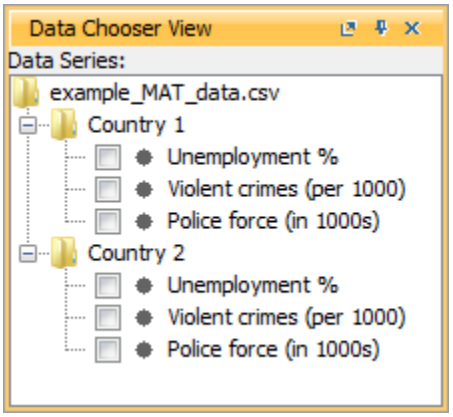
19 Click in the Country Name column, then click Add Data Category to add that column as a data category.

Create data categories to organize the data (that is, to create subcategories). Creating data categories displays your data in a hierarchical tree within the Data Chooser. The index of the row or column appears in the field, and the values of the row or column are shaded blue.

For more information on MAT's import options, see *Importing Data* on page 31.

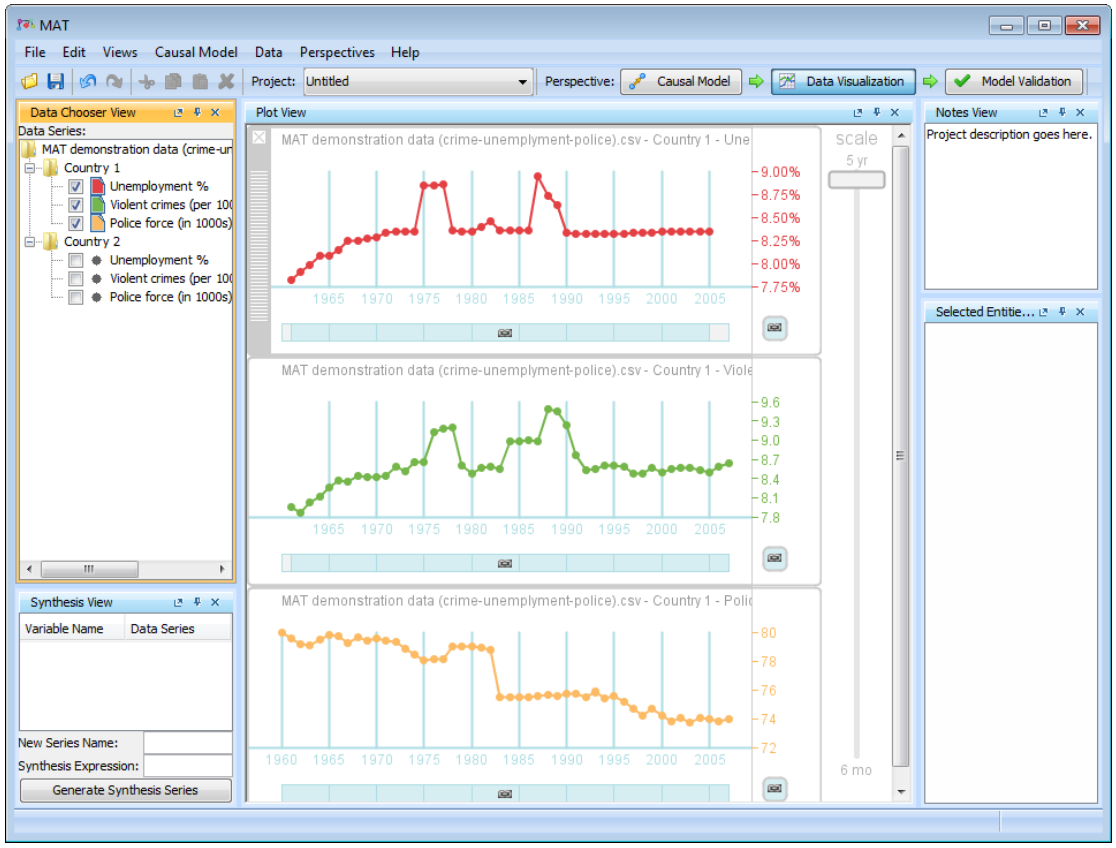
20 Click Import Data to display the data within the Data Chooser View.

Figure 8 Imported data displayed with data categories in the Data Chooser View



21 Check the boxes next to Country 1's Unemployment, Violent crimes, and Police force data series in the Data Chooser View to display the data series in the Plot View.

Figure 9 Three data series displayed in the Plot View



22 Drag the scale slider to the top (5 yr) to display the entire data series within the window.

For more information on the data visualizations available in MAT, see *Visualizing Data* on page 38.

23 Select File > Save to display the Save window.

24 Navigate to the directory where you want to save your project.

25 Enter CrimeAnalysis in the File name field and click Save to save the project as a new `.matprj` file.

For more information about MAT project files, see *Working with MAT Projects* on page 58.

Define data features

In this section of the tutorial, you will define features (or events) in the data.

26 Drag the mouse from 1974 to 1978 in the Unemployment data series graph in the Plot View to create a feature and display the Set Concept Type window.

Figure 10 Creating a data feature

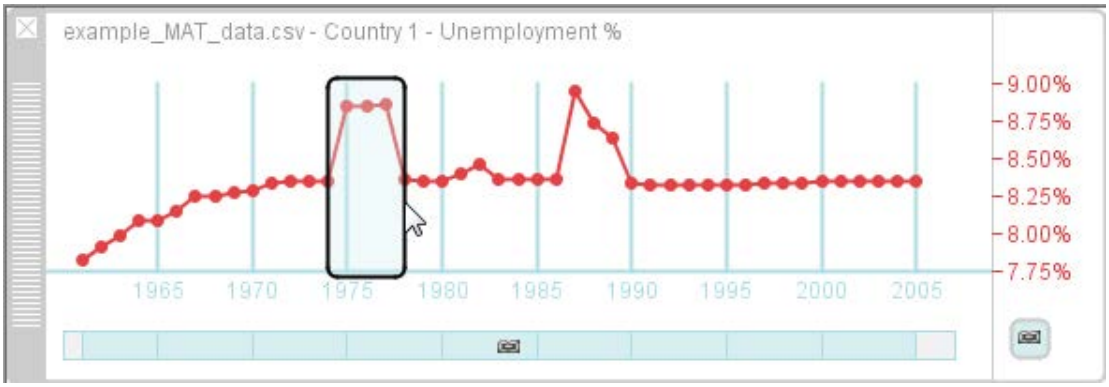
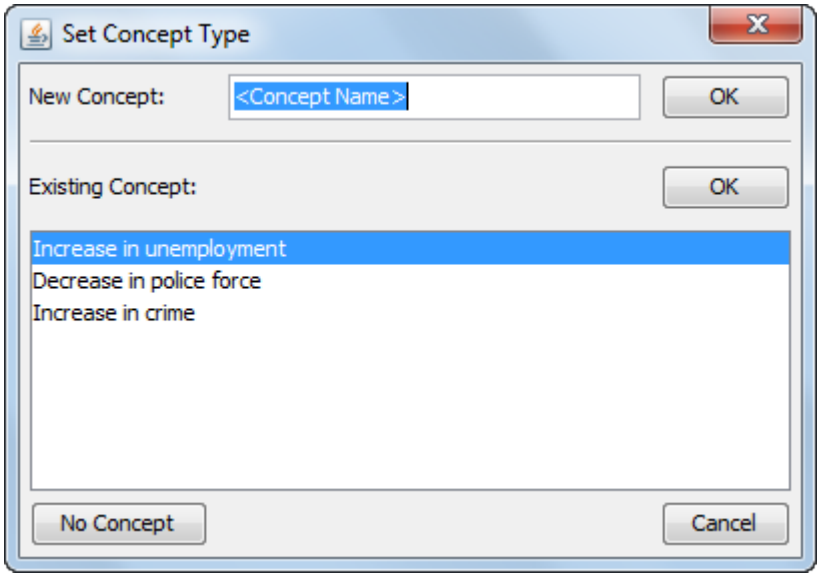


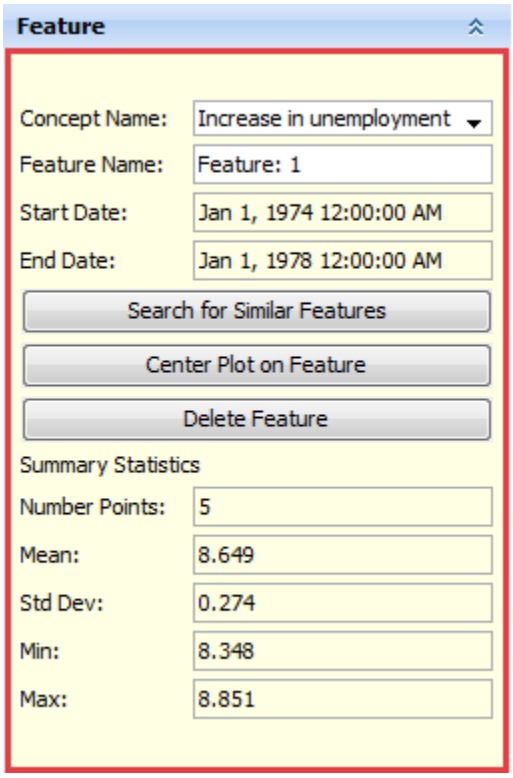
Figure 11 Associating the feature with a concept



27 Select the correct concept from the Existing Concept list (in this case, Increase in unemployment) and click OK in the Existing Concept area to create the data feature.

If you uncheck the data series, the dot next to the series is shown in red on the Data Chooser View to indicate that the data series contains one or more data features. The details of the feature are displayed in the Selected Entities View.

Figure 12 Data feature details



28 Repeat steps 26 and 27 for the second increase in unemployment, both increases in crime, and the decrease in police force in those data series, so the Plot View looks similar to Figure 13. (Be sure to select the correct concept from the list for each feature.)

Figure 13 Defined features in the data series

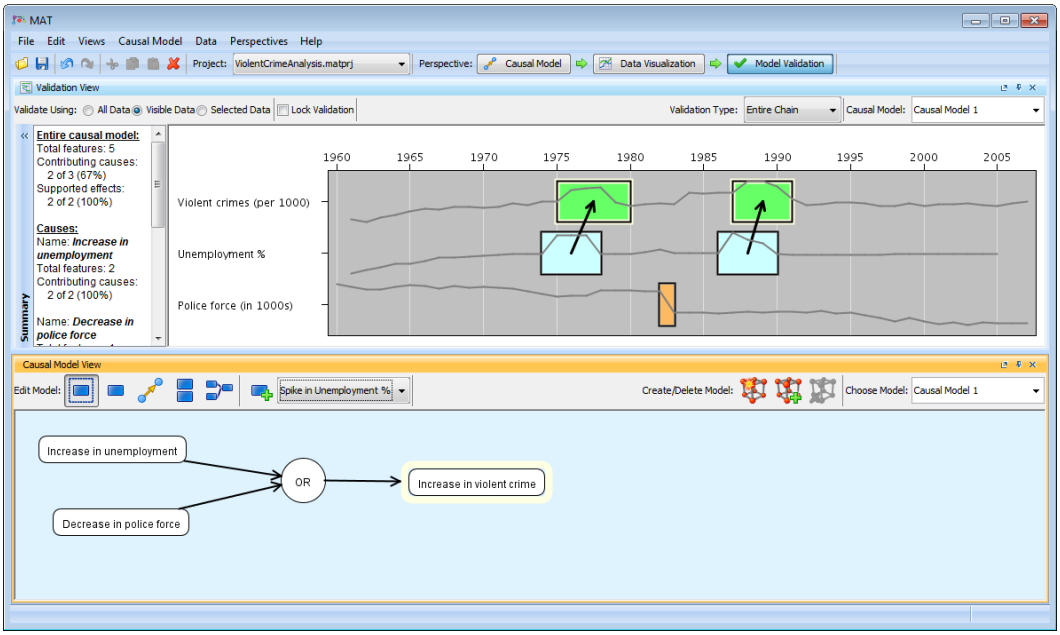
For more information, see *Working with data features* on page 48.

Validate the model

In this section of the tutorial, you will analyze the validity of your causal model.

29 Click Model Validation in the MAT toolbar to display the Model Validation perspective and validate the causal model.

Figure 14 Validating the causal model using the data displayed in the Plot View



30 Review the results in the Validation View.

The Validation View displays the features in the model in a timeline for the selected data. It also indicates whether each features is supported or contributes to the effect. The features you created are displayed in the following colors:

- Light blue – Contributing cause – Cause that directly supports an effect
- Orange – Non-contributing cause – Cause which does not contribute support to an effect
- Green – Supported effect

Click a node in the Causal Model View to highlight the features in the Validation View. Click a feature in the Validation View to display the causal model that contains the relevant concept in the Causal Model View and see the causal links in the Validation View. Mouse over the data series names in the Validation View to see the actual data series.

You can see that increased unemployment is a cause of increase in violent crimes. The decrease in police force does not appear to contribute to this effect within the parameters defined in the causal model.

4 WORKING WITH CAUSAL MODELS

This chapter describes how to create and edit causal models within the Model Analyst's Toolkit.

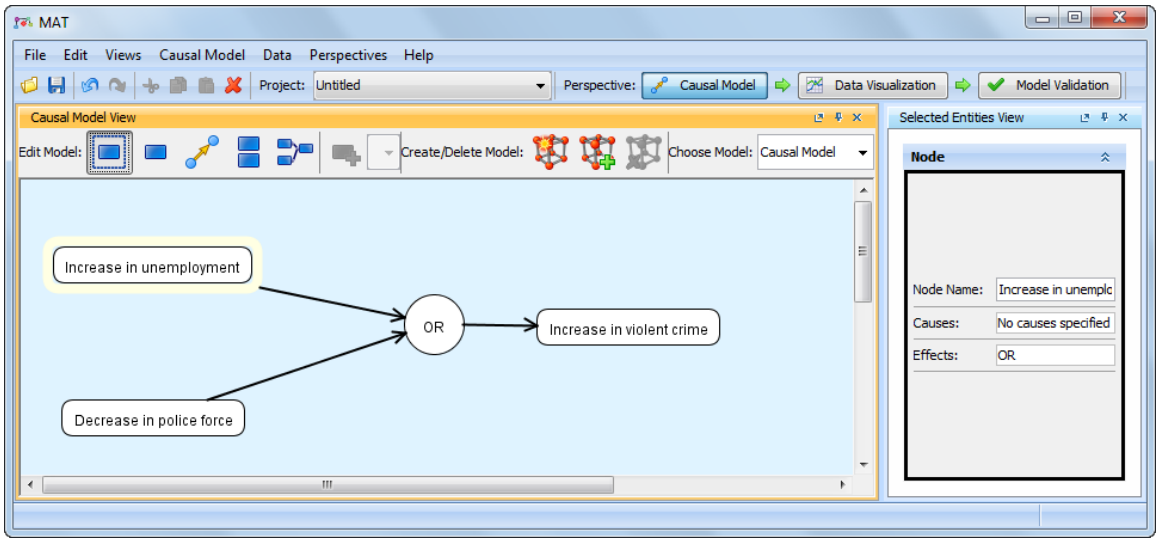
This chapter includes the following topics:

Overview

In MAT, a causal model represents any set of concepts with a cause-effect relationship. For example, you might theorize that lowering interest rates increases the valuation of the stock market, a simple two-concept model. Models can be of virtually anything; they can be social networks, economic theories, political science, physical events, or chemical reactions. Anything with a cause-effect relationship can be modeled in MAT.

Concepts are represented by *nodes*. Each node has a set of *data features* associated with it. Relationships between concepts are represented by *links*. A *causal model* is a collection of nodes and links that represents your theory of cause and effect.

Figure 15 Causal Model perspective



Creating a causal model

You can create a new causal model by drawing the model, copying an existing model, or MAT can generate a recommended models based on your imported data.

Creating a new causal model

To create a new causal model

1 Click  in the Causal Model View toolbar to create a new, empty causal model.

2 Click in the Choose Model drop-down list to rename the model.

3 Use the tools in the Causal Model View toolbar to create the nodes, groups, and links in the model.

For more information, see *Working with nodes* on page 26 and *Working with causal links* on page 28.

To copy an existing causal model

1 Click  in the Causal Model View toolbar to create a new causal model that is a copy of the model displayed in the Causal Model View.

The new model is displayed in the Choose Model drop-down list as *Causal Model* copy.

2 Select *Causal Model* copy from the Choose Model drop-down list.

3 Click in the Choose Model drop-down list to rename the model.

4 Use the tools in the Causal Model View toolbar to modify the nodes, groups, and links in the model.

For more information, see *Working with nodes* on page 26 and *Working with causal links* on page 28.

To rename a causal model

1 Select the causal model you want to rename from the Choose Model drop-down list.

2 Click in the Choose Model drop-down list and edit the name of the model.

To delete a causal model

1 Select the causal model you want to delete from the Choose Model drop-down list.

2 Click  in the Causal Model View toolbar to delete the model.

Generating recommended causal models from the data

MAT can generate causal models based on an analysis of your imported data. MAT recommends causal models included in a Pareto Frontier based on:

- Performance – Number of supported effects and contributing causes
- Model size – Number of nodes and edges
- Temporal aspects – Size of temporal window

MAT's recommendations include both simple causal models with a single cause for the effect, and more complex causal models with multiple causes combined using logic nodes.

You can explore the recommended causal models and see how they influence model validation. For example, the following two figures show how you can quickly switch between viewing a simple causal model with a larger temporal offset (seven years) and a more complex causal model with a smaller temporal offset (one year).

Figure 16 Viewing a simple recommended causal model with a seven-year temporal offset

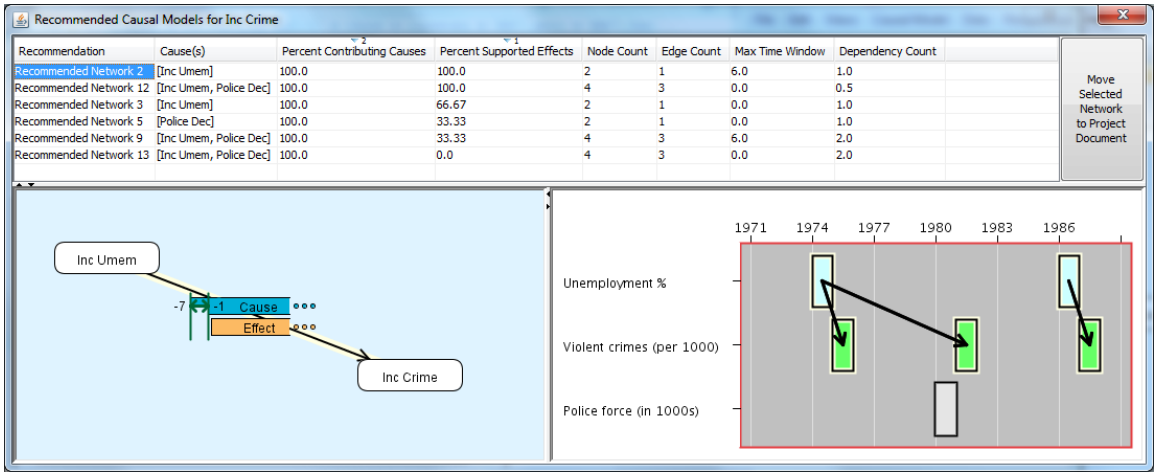
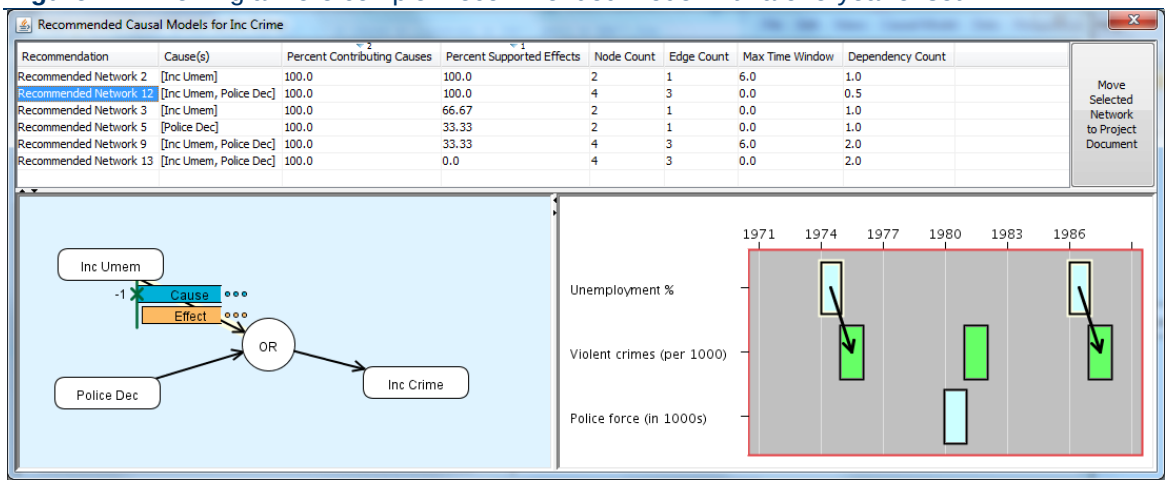
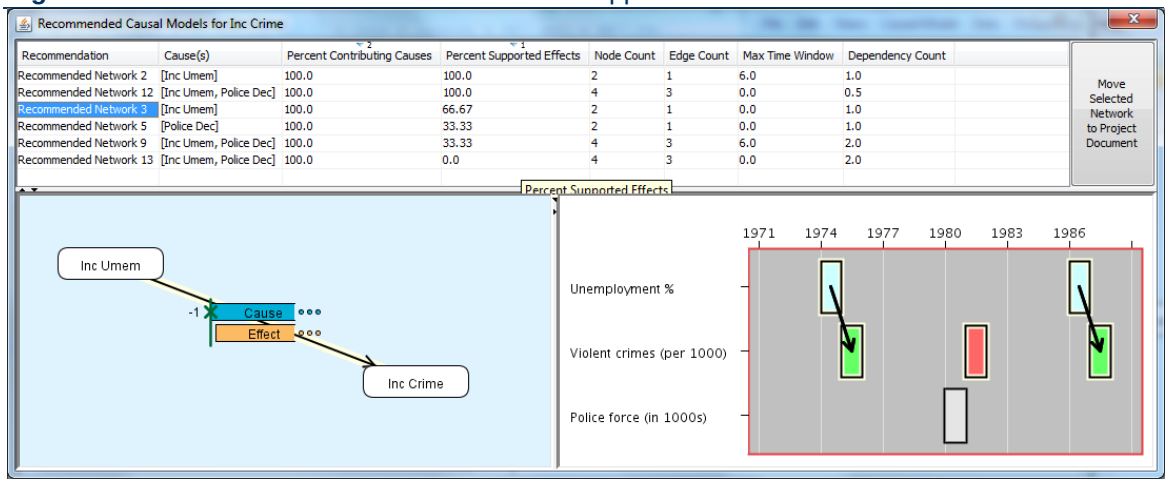


Figure 17 Viewing a more complex recommended model with a one-year offset



MAT also recommends causal models with lower performance, such as simple causal models with small temporal windows (as shown in Figure 18), even if they do not provide support for all of the effects.

Figure 18 Recommended model that does not support all effects



To generate recommended causal models from the data

1 Import your data.

For more information and instructions, see *Importing Data* on page 31.

2 Visualize your data and associate data features to your concepts.

For more information and instructions, see *Visualizing Data* on page 38 and *Working with data features* on page 48.

3 Right-click a node to configure your recommendation settings.

You can configure which features are used to generate the recommended causal models. You can select whether you want to use your existing features, automatically extract features from your imported data, or a combination of both.

Select:

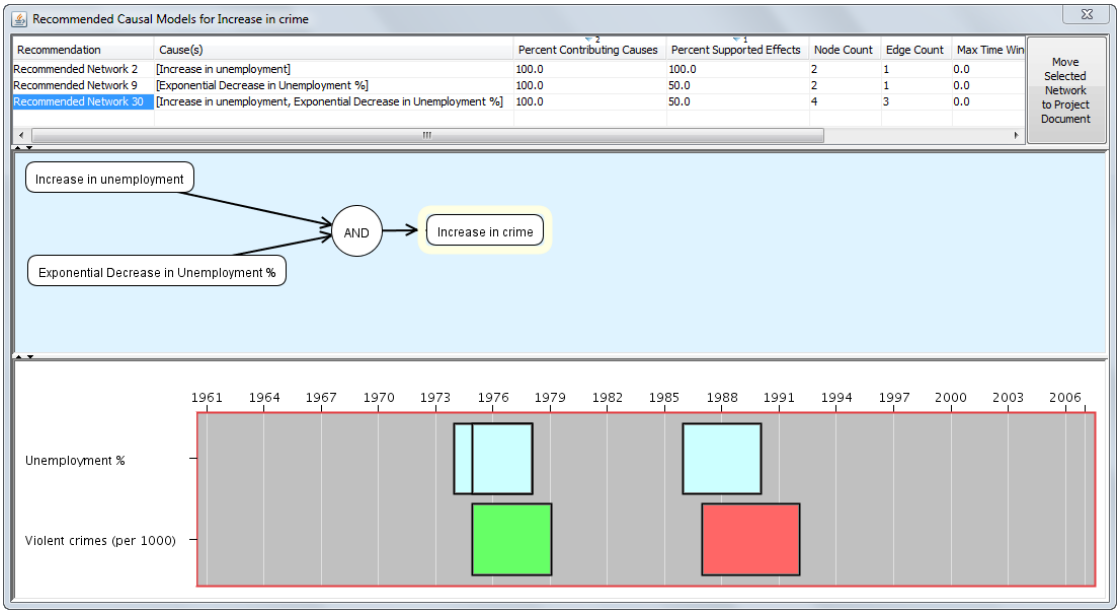
- Uses no existing features to use only extracted features
- Uses existing features from opened series to use features you defined on series that are displayed in the Data Visualization perspective
- Uses existing features from all series to use all the features you defined, whether or not they are defined on series displayed in the Data Visualization perspective
- Uses no extracted features to use only features you defined
- Uses extracted features from opened series to extract features from the series that are displayed in the Data Visualization perspective
- Uses extracted features from all series to extract features from all the series in your imported data

You can select one option each for existing features (that is, those you've defined) and extracted features.

4 Right-click a node and select:

- Recommend Causal Models Using Pruned Search of Existing/Extracted Features if you want MAT to quickly return recommendations. MAT first builds simple models, then uses those models to create more complex models. This option may cause some valid models to be missed.
- Recommend Causal Models Using Exhaustive Search of Existing Features if you want MAT to examine all possible combinations of causes with all possible combinations of temporal offsets. The option will return all valid models, but may take quite some time. (A progress bar is displayed and you can cancel this operation if it takes too long.)

Figure 19 Three causal models recommended by MAT based on the imported data



The table at the top of the window contains the causal models that MAT recommends from an analysis of the imported data. For each recommended causal model, the following information is displayed:

- Causes – List of potential causes for the selected node.
- Percent Contributing Causes – Percentage of cause features that are contributing to an effect.
- Percent Supported Effects – Percentage of effect features that are supported by a cause.
- Node Count – The number of nodes in the recommended model. Indicates model complexity.
- Edge Count – The number of edges in the recommended model. Indicates model complexity.
- Max Time Window – The longest time between onset of the cause and the onset of the effect.
- Max Time Offset – The longest time between the onset of the cause and onset of the effect.
- Dependency Count – Number of causes the effect is dependent upon. The higher the dependency count, the more restrictive the model. For example, if the model has two causes linked by an AND, it has two dependencies. If the causes are combined with an OR, it has 0.5 dependencies because it is not depending solely on either cause.

Click a column header to sort the table by that values in that column. Click again to reverse the sort.

5 Select the model you want to include in your project and click Add Selected Causal Model to Project to create a new causal model in the project and display it in the Choose Model drop-down list.

6 Edit the causal model if necessary.

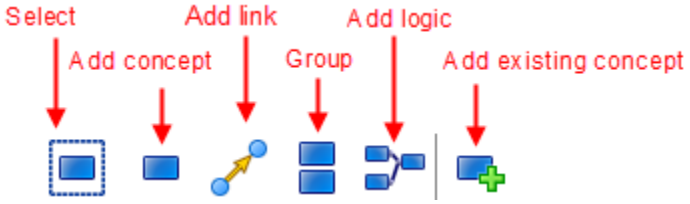
For more information, see *Working with nodes* on page 26 and *Working with causal links* on page 28.

Working with nodes

You can create concept and logic nodes, move nodes, group nodes, and rename them. Create a node to represent a concept. Create a group to associate similar concepts. A relationship with a group applies to all the concepts within the group. Logic nodes support complex relationships between nodes and groups.

You can use the Graph View toolbar to work with nodes and edges in your network.

Figure 20 Graph View toolbar



To create a concept node

7 Click the concept node tool or select from the existing concept drop-down list and click in the Causal Model View toolbar.

For existing concepts to appear, you must have defined features on your data series.

8 Click in the Causal Model View window to create the node.

9 Click the node again to rename it or edit the Node Name field in the Selected Entities View.

To create a logic node

1 Click the logic node tool in the Causal Model View toolbar.

2 Click in the Causal Model View window to create the logic node.

3 Select the type of logic to use from the Logic Node Type drop-down list in the Selected Entities View.

Select:

- AND if all nodes connected to the AND node must be true for a cause-effect relationship
- OR if just one of the nodes connected to the OR node must be true for a cause-effect relationship
- NOT if the node connected to the NOT node must be false for a cause-effect relationship

To move a node

4 Click the select tool in the Causal Model View toolbar.

5 Drag a node to move it.

To group nodes

1 Click the group concept nodes tool in the Causal Model View toolbar.

2 Click in the Causal Model View to create a group.

3 Drag a node into the group to add it to the group. Drag a node out of the group to remove it from the group.

MAT treats the nodes within the group as if they are connected by a logical OR.

To rename a node or group

1 Click the node or group you want to rename to select it.

2 Click the node or group again to edit its name or edit its name in the Selected Entities View.

If you rename a logic node to AND, OR, or NOT, the node's logic changes to match the name.

To delete a node or group

1 Click the node or group to select it.

2 Click on the MAT toolbar, press Delete, or select Edit > Delete from the MAT menubar.

Working with causal links

Create a causal link to describe a causal relationship between nodes and groups.

To create a causal link

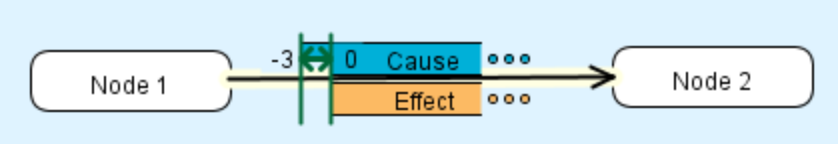
3 Click the causal link tool in the Causal Model View toolbar.

4 Click a node or group.

5 Click another node or group to link the two elements.

The link is drawn from the first node to the second node and a graphical display of constraints is displayed, unless the link originates from a logic node.

Figure 21 Graphical display of link constraints



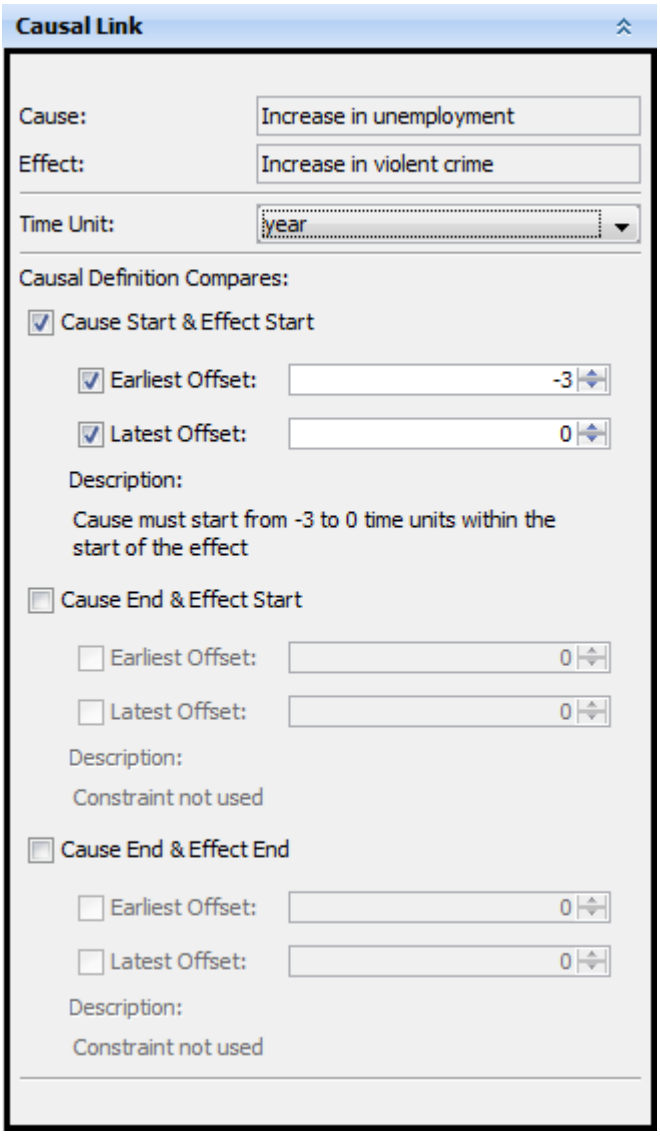
For example, in this figure, the causal definition compares the start of the cause (Node 1) and the start of the effect (Node 2), with an offset of -3 time units. That is, when MAT validates the model, features are considered causative if the start of the feature associated with Node 2 occurs within 3 time units of the start of the feature associated with Node 1.

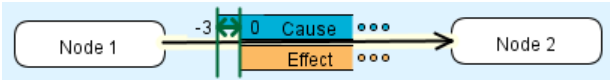
6 Set time constraints on the relationship in the Selected Entities View.

Constraints specify the time range within which a feature must occur to be considered a cause or effect. Once defined, you can adjust the offset by dragging the offset bar in the link constraint graphic.

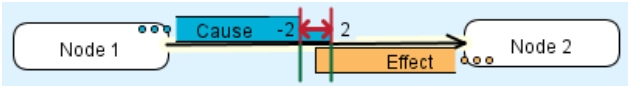
7 Enter or review the following attributes of the link:

- Cause – Displays the name of the causal node.
- Effect – Displays the name of the effect node.
- Time Unit – Select from the drop-down list to specify the time units for the offsets used to determine causation.
- Check any combination of the Causal Definition Compares boxes to specify the time range for causality. The constraints are shown on the link constraint graphic within the Causal Model View.
- Cause Start & Effect Start – Check this box to compare the start of a data feature to the start of the effect to determine causation. Enter the greatest number of time units you want to consider causative in the Earliest Offset box. That is, a causative data feature must start within this number of time units before the effect data feature starts. Use a negative number to specify the number of units before the start of the effect. Enter zero if a data feature must begin in the same time unit as the effect to be considered a cause. Enter the other end of the time range (usually zero) in the Latest Offset box. The description changes based on the units you select.

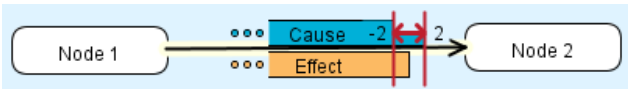




- Cause End & Effect Start – Check this box to compare the end of a data feature to the start of the effect to determine causation. Enter units in the Earliest Offset and Latest Offset boxes to specify the time range before the effect starts in which a data feature must end to be considered causative.



- Cause End & Effect End – Check this box to compare the end of a data feature to the end of the effect to determine causation. Enter units in the Earliest Offset and Latest Offset boxes to specify the time range before the effect ends in which a data feature must end to be considered causative.



To delete a causal link

1 Click the link to select it.

2 Click  on the MAT toolbar, press **Delete**, or select **Edit > Delete** from the MAT menubar.

5 IMPORTING DATA

You can import any text-based dataset into the Model Analyst's Toolkit, or use the provided sample data. MAT can handle files generated by Excel, text editors, SAS, or SPSS.

This chapter includes the following topics:

- Importing your data
- Configuring data for import
- Saving a configuration for a data file
- Importing default data

Importing your data

Any text-based data can be imported into the Model Analyst's Toolkit. MAT can handle files generated by Excel, text editors, SAS, or SPSS.

To import data into the Model Analyst's Toolkit

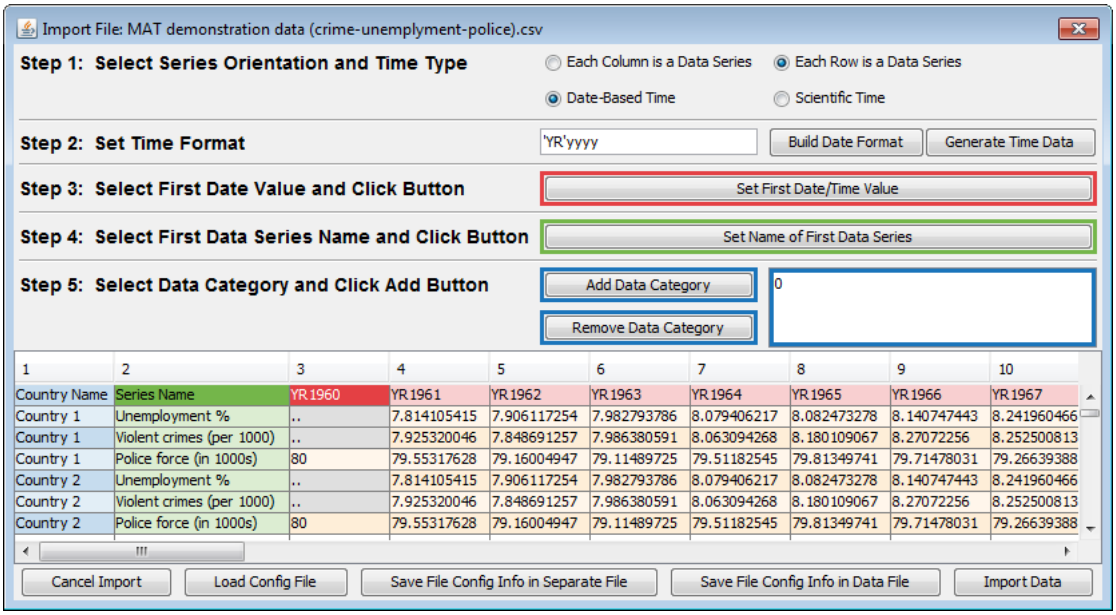
1 Click Data Visualization in the MAT toolbar or select Perspectives > Data Visualization from the menubar to display the Data Visualization perspective.

2 Select File > Import Data from the menubar to display the Open window.

3 Navigate to the directory that contains your dataset, select the data file, and click Open. You can also double-click the data file to open it.

If you previously imported the file and saved the file configuration within the data file, MAT imports the data automatically. If you did not save the configuration in the data file, MAT displays the Import File window.

Figure 22 Importing a dataset containing unemployment, crime, and police force statistics for two countries



1	2	3	4	5	6	7	8	9	10
Country Name	Series Name	YR1960	YR1961	YR1962	YR1963	YR1964	YR1965	YR1966	YR1967
Country 1	Unemployment %	..	7.814105415	7.906117254	7.982793786	8.079406217	8.082473278	8.140747443	8.241960466
Country 1	Violent crimes (per 1000)	..	7.925320046	7.848691257	7.986380591	8.063094268	8.180109067	8.27072256	8.252500813
Country 1	Police force (in 1000s)	80	79.55317628	79.16004947	79.11489725	79.51182545	79.81349741	79.71478031	79.26639388
Country 2	Unemployment %	..	7.814105415	7.906117254	7.982793786	8.079406217	8.082473278	8.140747443	8.241960466
Country 2	Violent crimes (per 1000)	..	7.925320046	7.848691257	7.986380591	8.063094268	8.180109067	8.27072256	8.252500813
Country 2	Police force (in 1000s)	80	79.55317628	79.16004947	79.11489725	79.51182545	79.81349741	79.71478031	79.26639388

4 Load a configuration file or create a new configuration for the data.

For more information, see *Configuring data for import* on page 33.

5 (Optional) Save the import configuration.

Click:

- Save File Config Info in Separate File to save the configuration information in a separate file.

- Save File Config Info in Data File to save the configuration information in the `.csv` data file.

For more information, see *Saving a configuration for a data file* on page 35.

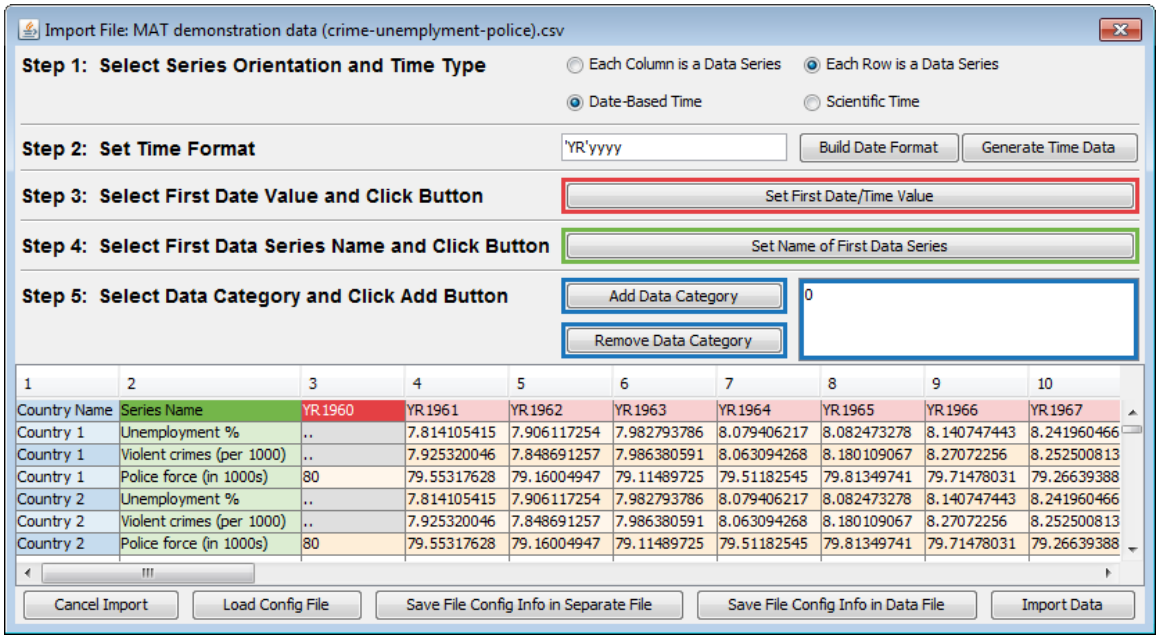
6 Click Import Data to import the data into MAT and display it in the Data Chooser View.

For more information, see *Visualizing Data* on page 38.

Configuring data for import

MAT can help you analyze both date-based data (such as the demographic data) and scientific time data (such as data from a neurological experiment). Date-based data uses a calendar, while scientific time is relative to a starting time. You must provide MAT with a few cues so that it can correctly import your data. We call these cues a configuration, and you enter them on the Import File window.

Figure 23 Providing configuration information on a data file



1	2	3	4	5	6	7	8	9	10
Country Name	Series Name	YR1960	YR1961	YR1962	YR1963	YR1964	YR1965	YR1966	YR1967
Country 1	Unemployment %	..	7.814105415	7.906117254	7.982793786	8.079406217	8.082473278	8.140747443	8.241960466
Country 1	Violent crimes (per 1000)	..	7.925320046	7.848691257	7.986380591	8.063094268	8.180109067	8.27072256	8.252500813
Country 1	Police force (in 1000s)	80	79.55317628	79.16004947	79.11489725	79.51182545	79.81349741	79.71478031	79.26639388
Country 2	Unemployment %	..	7.814105415	7.906117254	7.982793786	8.079406217	8.082473278	8.140747443	8.241960466
Country 2	Violent crimes (per 1000)	..	7.925320046	7.848691257	7.986380591	8.063094268	8.180109067	8.27072256	8.252500813
Country 2	Police force (in 1000s)	80	79.55317628	79.16004947	79.11489725	79.51182545	79.81349741	79.71478031	79.26639388

To configure data for import

7 On the Import File window, select the series orientation. Select:

- Each Column is a Data Series if values for each point in time for the series appear in a single column
- Each Row is a Data Series if values for each point in time for the series appear in a single row, as shown in the example in Figure 23.

8 Select the time format: Select:

- Date-Based Time if your time values are based on calendar dates

- Scientific Time if your time values are time offsets from some starting time (or if your data has a higher resolution than a millisecond *(Scientific time is not supported in this release of the Model Analyst's Toolkit)*

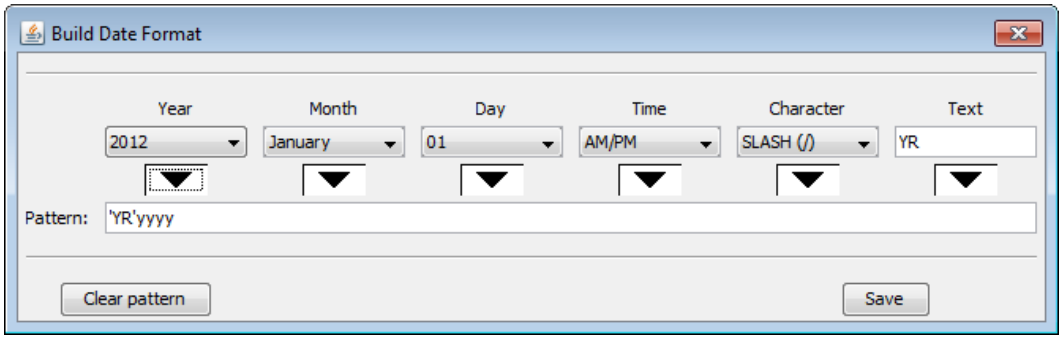
9 Specify the format in which your time data appears in one of the following ways:

- Enter the format in the field. Some common formats are yyyy (for example, 2014), mm/dd/yy (03/15/14), or MMM d, yyyy (Jan 1, 2014).

  Possible codes include: `yyyy`, `yy`, `MMMM`, `MMM`, `MM`, `M`, `dd`, `d`, `a`, `HH`, `H`, `hh`, `h`, `mm`, `m`, `ss`, `s`, `SS`, `S`, slash, space, dash, comma, period, and 'other' (any characters contained within single quotes). Examples of each code appear within the drop-down lists.

  Click Build Date Format to display the Build Date Format window. This window provides all the codes above and examples to help you correctly describe the date format.

Figure 24 Building a date format

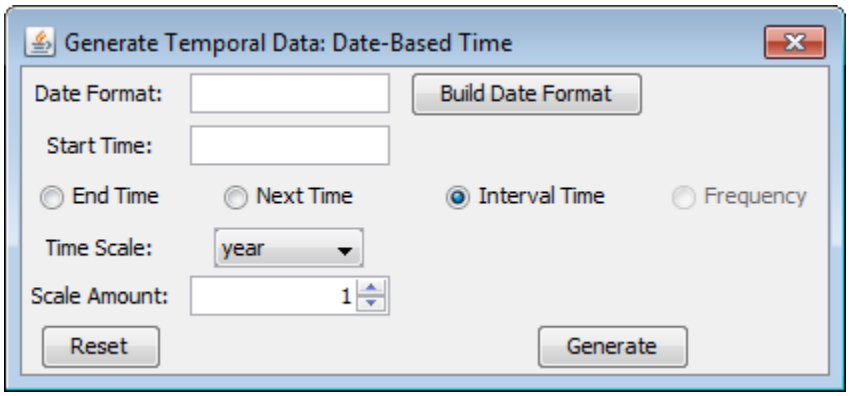


Select from each drop-down and click the down-arrow to enter the code for the selected example into the Pattern field.

Click Clear Pattern to clear the Pattern field. Click Save to save the format and display it on the Import File window.

- Click Generate Time Data to display the Generate Temporal Data window.

Figure 25 Generating time data



Enter information into the following elements on the window:

- Date Format – Enter the format you want to use for the date or click **Build Date Format** to display the Build Date Format window.
- Start Time – Enter the starting time for the data
- Select one of the following radio buttons to specify the time scale. Select:
  - End Time to enter just the last time data was collected
  - Next Time to enter the time of the next data point
  - Interval Time to enter the time between data points
  - Frequency to enter the frequency with which the data was recorded
- Time Scale – Select the units of time from the drop-down list.
- Scale Amount – Enter the number of time units into the field. For example, if you want to create a new date every two years, enter 2 here, and select year from the Time Scale drop-down list.

Click **Reset** to clear the elements in the window. Click **Generate** to generate the date/time values for the data file.

10 Click the cell in the data table that displays the first time value, then click **Set First Date/Time Value**.

The selected cell is shaded red. If the date/time format matches the format you specified, the other date/time values in your dataset are shaded pink. Correct the format until the values are correctly shaded, or MAT will not be able to import your data.

11 Click in the cell in the data table that displays the name of the first data series, then click **Set Name of First Data Series**.

The selected cell is shaded green. All data series in your dataset are shaded light green. Date/time values for each data series in the data table are shaded yellow and gray. Gray shading indicates a missing value.

12 (Optional) Create data categories to organize the data (that is, to create subcategories). Creating data categories displays your data in a hierarchical tree within the Data Chooser.

Select a row or column header, then click **Add Data Category** to add that column as a data category. The index of the row or column appears in the field, and the values of the row or column are shaded blue. Select the index in the field and click **Remove Data Category** to remove that category.

Saving a configuration for a data file

Once you provide information to MAT about a data file, you can choose to save that configuration data in a separate file or within the data file.

Configuration data is saved as a line of comma-separated values, prefaced by a pound sign (#). When you import a data file into MAT, any lines beginning with # are treated as file format metadata. Data files with all the required metadata contained inside the file are imported automatically. You do not need to edit any data file configuration values manually.

MAT saves the following attributes of the data file format at the beginning of the data file (example values are shown):

```
#,dataSeriesOrientation,ROW
#,dateFormatString,'YR'yyyy
#,firstDateRow,0
#,firstDateCol,2
#,firstDataSeriesNameRow,1
#,firstDataSeriesNameCol,1
#,dataCategories,0
#,hasGeneratedTime,false
#,timeData
```

Variable	Description
`dataSeriesOrientation`	Has two possible values: `ROW` or `COLUMN`.
`dateFormatString`	Java Date pattern that allows the date values in the file to be parsed.
`firstDateRow`	Row in the data file that contains the first date value. Row indices start at zero.
`firstDateCol`	Column in the data file that contains the first date value. Column indices start at zero.
`firstDataSeriesNameRow`	Row in the data file that contains the name of the first data series.
`firstDataSeriesNameCol`	Column in the data file that contains the name of the first data series.
`dataCategories`	List of rows or columns that contain attributes for each of the data series. If the data series are oriented in columns, this variable contains a list of rows. If the data series are oriented in rows, this variable contains a list of columns. Data categories are used to organize the data within MAT.
`hasGeneratedTime`	Displays `TRUE` if MAT generated date/time values, `FALSE` if no values were generated.
`timeData`	The date/time values generated by MAT.

Importing default data

The Model Analyst's Toolkit includes a set of configured, imported data from the `example_MAT_data.csv` file located in the `MAT7.0.0\data` directory. You can change this default dataset to another dataset that you have configured for import.

To import the default data

Select File > Import Default Data from the menubar to import the default data file and display it in the Data Chooser View.

To change the default dataset

1 Configure the dataset you want to be the default dataset. Once you have configured the data, click Save File Config Info in Data File to save the configuration information within the `.csv` data file.

For more information, see *Configuring data for import* on page 33 and *Saving a configuration for a data file* on page 35.

2 Select File > Choose Default Data File from the menubar to display the Open window.

3 Navigate to the directory that contains your data file, select it, and click Open to import and display the dataset in the Data Chooser View. You can also double-click the data file to open it.

When you select File > Import Default Data from the menubar, the selected data file will now open.

If you did not correctly configure the data, it will not display in the Data Chooser View. If this occurs, make sure you configured the data correctly.

6 VISUALIZING DATA

MAT offers a number of visualizations for your data and allows you to identify and define data features within your data series.

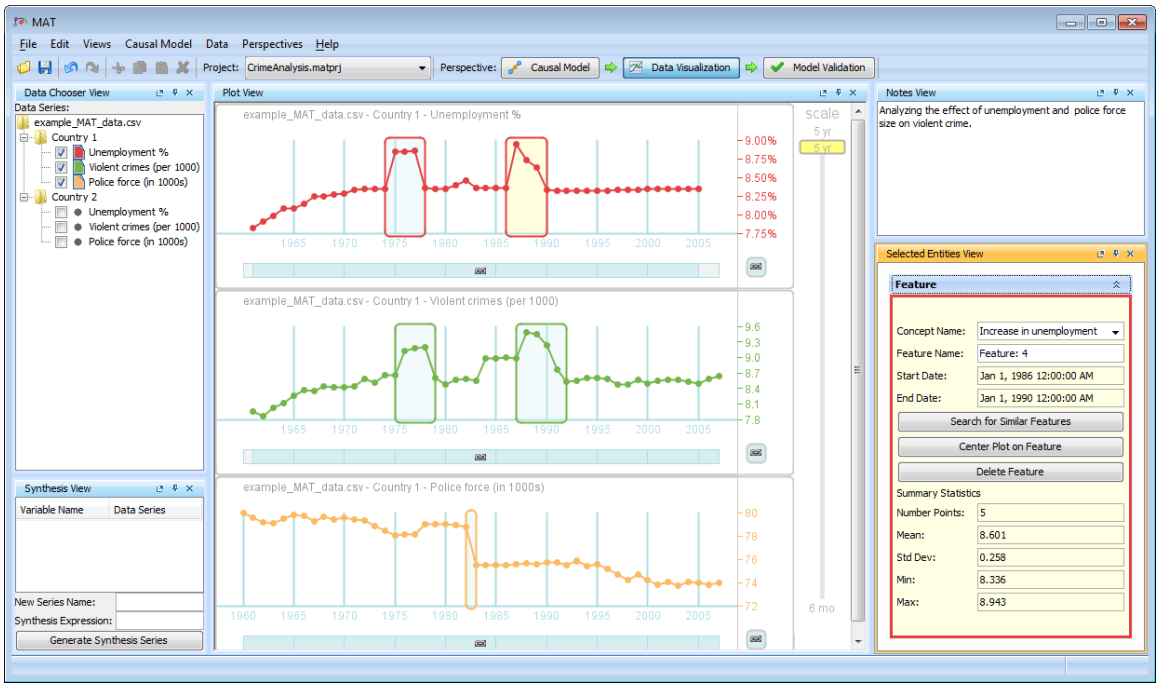
This chapter includes the following topics:

- Overview
- Selecting a data series
- Exploring data with the Plot View
- Visualizing multiple data series
- Working with data features

Overview

The Data Visualization perspective helps you explore the data you imported into MAT. This data can be correlated and examined through different statistics and different time period offsets. This perspective can also be used to define features and associate features with the concepts in your causal model.

Figure 26 Data Visualization perspective



To visualize your imported data

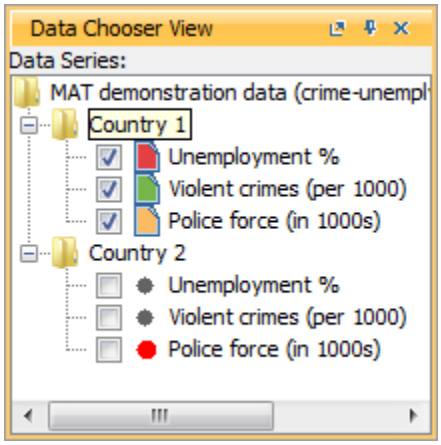
1 Click Data Visualization in the MAT toolbar or select Perspectives > Data Visualization from the menubar to display the Data Visualization perspective.

2 Check the box next to a data series in the Data Chooser View to display that series in the Plot View.

Selecting a data series for display

Data series from imported datasets are displayed in the Data Chooser View. This view allows you to select data series for display in the Plot View, search for a data series by name, and remove data series.

Figure 27 Imported data showing nested categories



Double-click the data file (the top-level folder) to open or close it.

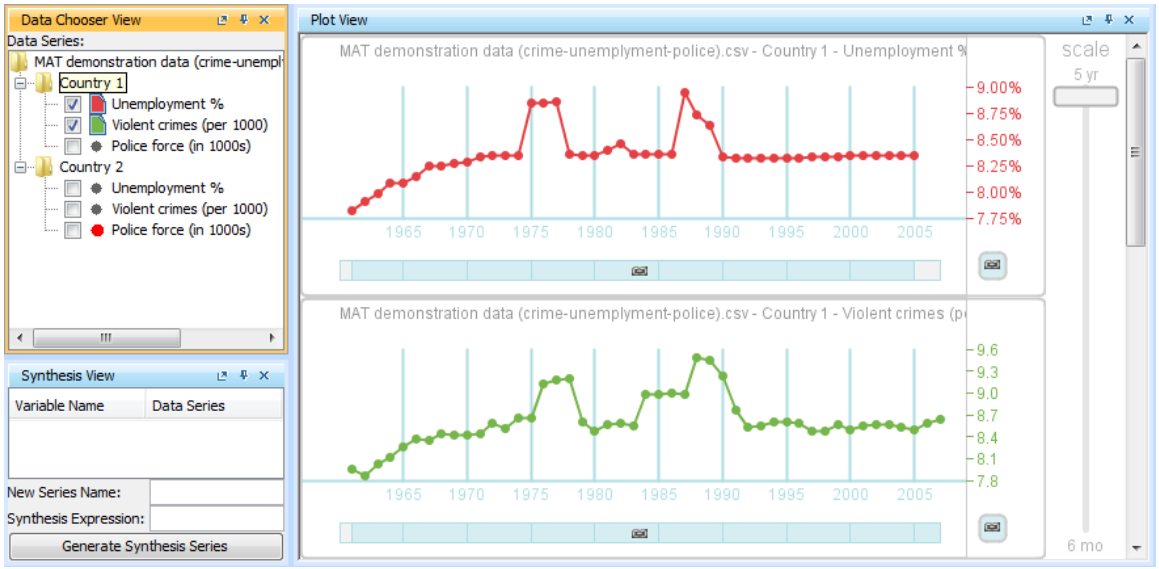
If you defined categories when you imported your file, your data file appears as a series of nested folders. Click the + next to a folder to open it. Click – to close it.

A red dot (●) next to a series indicates that features have been defined for the data series. Colored page icons appear next to a series that is currently displayed in the Plot View.

To display a data series

Check the box next to a data series to display it in the Plot View. Uncheck the box to remove it from the Plot View.

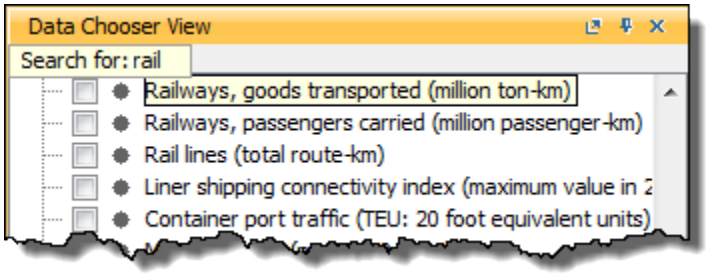
There is no limit to the number of data series you can display. However, displaying a large number of data series may slow performance.



To search for a data series

Click anywhere within the Data Chooser View and enter your search term. As you type, your term appears at the upper left of the Data Chooser View, and the first data series name that matches your term is selected. Use the up and down arrow keys to navigate through the results.

Figure 28 Searching for data series that begin with "rail"



You can use an asterisk (*) as a wildcard character. For example, enter *ship* to find data series where "ship" appears anywhere in the name. If you do not enter an asterisk as the first character in your search, MAT searches for data series that begin with your search term.

You can also use a vertical line (|) between search terms as a logical "or." This feature lets you search for multiple terms. For example, enter *women*|*female* to find the data series that have either "women" or "female" in the name.

To remove a data series from the Data Chooser View

Select a data series or data file and press Delete or click on the MAT toolbar to remove the data series or entire data file. You can delete multiple data series or files. Shift+click or Ctrl+click to select multiple data series or files for deletion.

The data is not deleted. If you remove a data series or file by mistake, select Edit > Undo from the MAT menubar or click in the toolbar. You can also re-import the data file.

Exploring data with the Plot View

When you display a data series in the Plot View, you can display detailed values, change the scale and color of the graph, link scrolling on multiple graphs to better visualize features offset in time, and combine multiple graphs to create one "stacked" graph.

Figure 29 Data series shown at a two year scale with linked scrolling

Mouse over a data point to display its value. The slider at the bottom shows you how much of your data is currently displayed. Use this slider to scroll your data. All graphs with linked scrolling will also scroll.

To change the scale

Use the scale slider to change the scale of all the graphs shown in the Plot View. The sliders at the bottom of each graph change to reflect the amount of data currently displayed.

To link scrolling on multiple charts

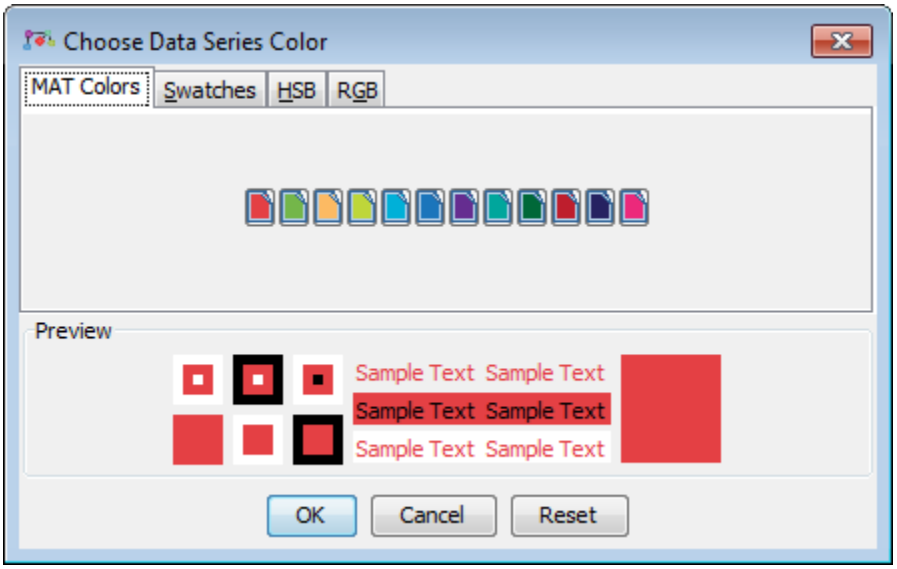
When multiple data series are displayed, you can link scrolling by clicking  . Click again to unlink scrolling.

You can use linked scrolling to view data series with features that are offset in time. Unlink the graphs, adjust one forward or backward in time with the scroll bar, then relink the graphs to link scrolling with the offset.

To change the color

1 Right-click the graph in the Plot View and select Change Color from the context menu to display the Choose Data Series Color window.

Figure 30 Selecting the color for a data series graph



2 Click a tab to select a color from different palettes. Click:

- MAT Colors to select a color from the MAT palette.
- Swatches to select a color from a common set of colors. As you click, colors are added to the Recent swatches.
- HSB to use hue, saturation, and brightness sliders to select a color.
- RGB to use red, , green, and blue sliders to select a color.

3 Select or create your color and click OK to display the graph in the selected color.

To stack charts

Drag a chart on top of another chart in the Plot View or select multiple charts, right-click and select Stack Charts from the context menu to combine the data on the charts.

Mouse over a point on the graph to display the name of the data series and the y axis for that series.

Click ⟶ to combine the y axes; click ⤖ to separate the axes.

Figure 31 Stacking percent unemployment on number of violent crimes with separate axes

To unstack a chart, right-click the stacked chart and select Unstack Data Series from the context menu.

To close a chart

Mouse over a chart and click ☒ at the upper left of the chart to remove it from the Plot View.

Creating a synthetic data series

You can create a new data series by applying an expression to existing data series. You can transform a single data series or combine multiple data series using mathematical expressions.

To create a synthetic data series

1. Drag one or more data series from the Data Chooser View to the Synthesis View.

   Each series is given a variable name.

2. Enter a name for the synthesized series in the New Series Name field.

3. Enter a synthesis expression using the variable names, in the Synthesis Expression field.

   You can create expressions using addition, subtraction, multiplication, division, log base 10, and natural log functions. For example: x1 + x2, x1 - x2, x1 * x2, x1/x2 , log(x1), ln(x1).

4. Click Generate Synthesis Series.

Visualizing multiple data series

You can display any of the following graphs for two selected data series:

- Scatterplot – Shows the relationship between two data series and uses linear regression to show the line that best fits the data
- Correlation offset plot – Displays the correlation between two data series using all possible time offsets
- Correlation matrix – Displays the correlation, significance, and number of values for two or more data series
- Dynamic time warp – Warps the timing of one data series to more closely match the shape of another series, where the amount of temporal warping can vary within specified limits; captures the effects that do not always follow causes by a constant amount of time

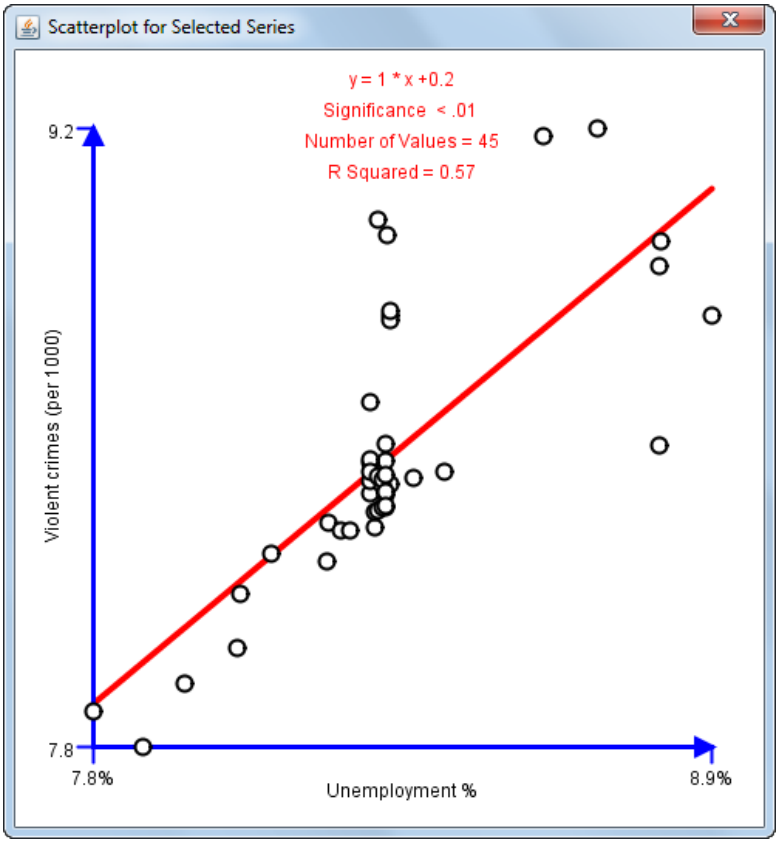
Each visualization recognizes and displays the correct units based on the selected data series.

To display a scatterplot

1. Ctrl+click to select two graphs in the Data Chooser View or Plot View.

2. Select Data > Display Scatterplot from the menubar or right-click one of the selected series and select Display Scatterplot from the context menu.

Figure 32 Scatterplot of unemployment and violent crimes

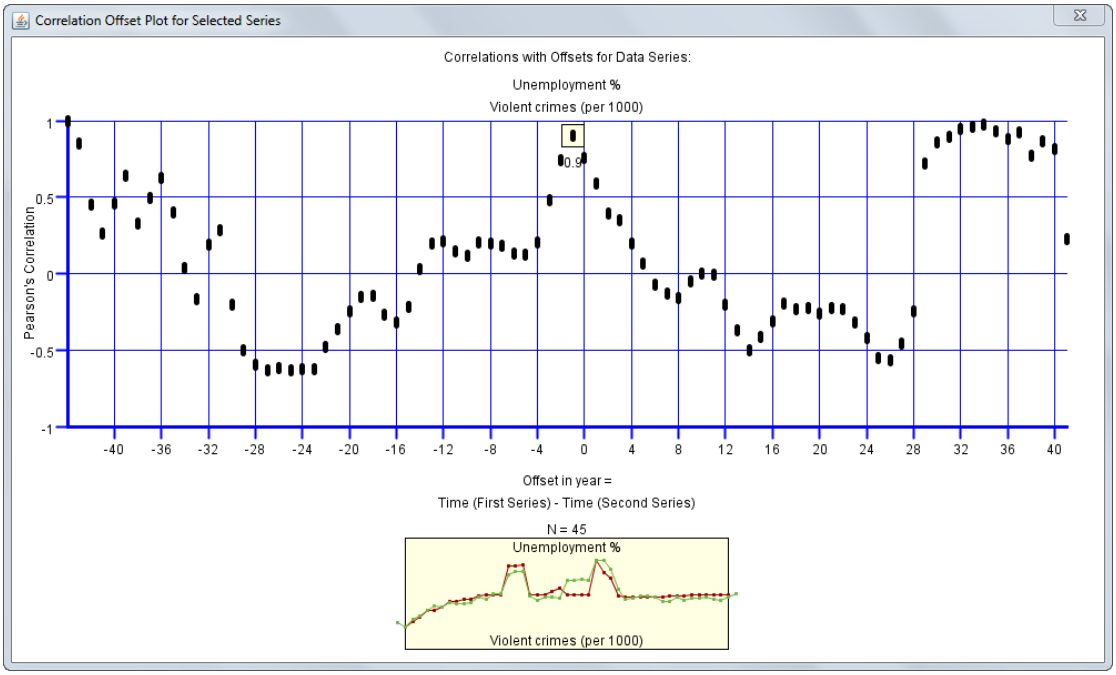


To display a correlation offset plot

1 Ctrl+click to select two graphs in the Data Chooser View or Plot View.

2 Select Data > Display Correlation Offset Plot from the menubar or right-click one of the selected series in the Plot View and select Display Correlation Offset Plot from the context menu to display the Correlation Offset Plot for Selected Series window.

Figure 33 Correlation offset plot of unemployment and violent crimes



3 Slide your mouse over the main graph to show the two graphs offset by that time value at the bottom of the window.

To display a correlation matrix

1 Ctrl+click to select two graphs in the Data Chooser View or Plot View.

2 Select Data > Display Correlation Matrix from the menubar or right-click one of the selected series in the Plot View and select Display Correlation Matrix from the context menu.

Figure 34 Correlation matrix for unemployment and violent crime

Correlation Matrix of Selected Series

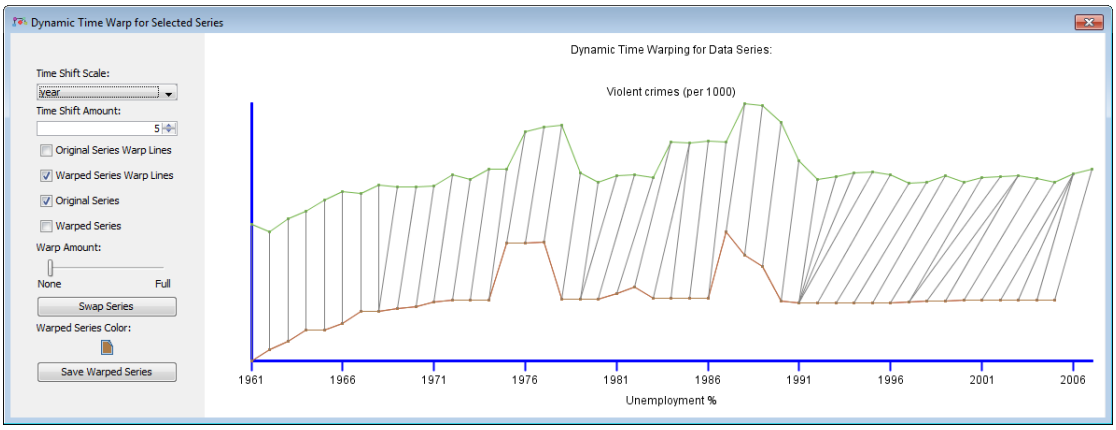
Correlation *Significance* *Number of Values*	(1)	(2)
(1) Unemployment %	1 0 45	
(2) Violent crimes (per 1000)	0.75 < 0.01 45	1 0 47

To display a dynamic time warp

1 Ctrl+click to select two graphs in the Data Chooser View or Plot View.

2 Select Data > Dynamic Time Warp from the menubar or right-click one of the selected series in the Plot View and select Dynamic Time Warp from the context menu to display the Dynamic Time Warp for Selected Series window.

Figure 35 Dynamic time warp for violent crime and unemployment



3 Specify the following parameters:

- Select a value from the Time Shift Scale drop-down list to set the units of time.
- Enter a number in the Time Shift Amount field to limit the amount of time that a data point can be shifted during warping.
- Check Original Series Warp Lines to display lines on the graph that link the points in the unwarped version of the data series to the points in the static data series that they will be warped to. This visualization makes sense only when you also check the Original Series box.
- Check Warped Series Warp Lines to display vertical lines on the graph that link the points in the warped version of the data series to the points in the static data series that they have been warped to. This visualization makes sense only when you also check the Warped Series box.
- Check Original Series to view the original, unwarped data at the bottom of the graph.
- Check Warped Series to view the newly warped data at the bottom of the graph. (You can view both the original and the warped series.)
- Use the Warp Amount slider to animate the warp from the original data to the warped data to highlight where warping is occurring.
- Click Swap Series to switch which of the two data series remains stable and which is warped. When visualizing a dynamic time warp, the hypothesized effect series should be on the top and the cause/predictor series should be on the bottom. If the cause/effect relationship is not clear, you can use this button to view how each data series must to be warped in time to align with the other series. Because warping always happens forward in time, the ordering matters and the cause (which happens first in time) should be the warped series.
- Click the Warped Series Color icon to display the Choose Data Series Color window. Select a color and click OK to change the color of the warped series. For more information, see *Exploring data with the Plot View* on page 41.

- Click Save Warped Series to display the Save New Data Series window. Enter the name, data category, and edit the name of the data file (if necessary) and click Save to save your warped series with your data set.

Working with data features

In MAT, you select the data series you want to use to test your theory and associate it with concepts in your model. The associations are between *features* of the data—which are subsets of the time series that have a distinctive characteristic, such as an increase or decrease, or an amount over a threshold—and concept *nodes* in your model.

You can create data features, edit them, and ask MAT to automatically recognize and recommend features in a data series.

Creating a data feature

You can create a data feature within any data series displayed in the Plot View.

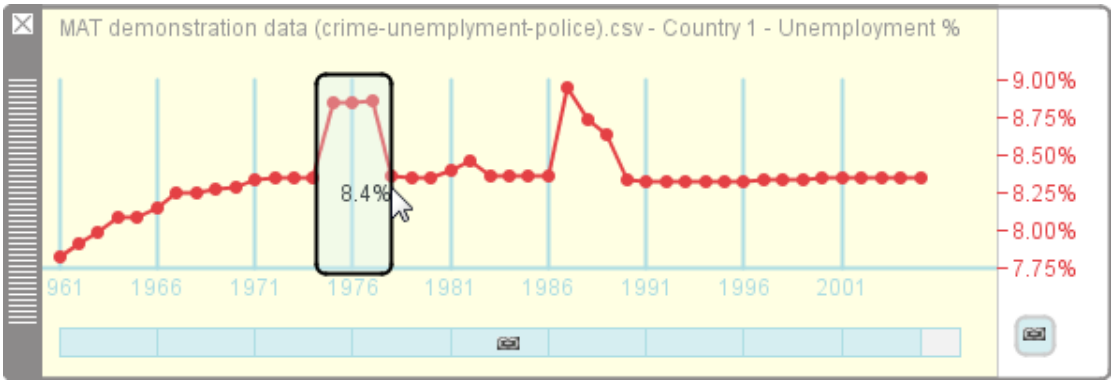
To create a data feature

1 Check a data series box in the Data Chooser View to display that series in the Plot View.

2 Adjust the scale slider so that the features in the data are clearly expressed in the graph.

3 Drag your mouse over the graph to create a feature and display the Set Concept Type window.

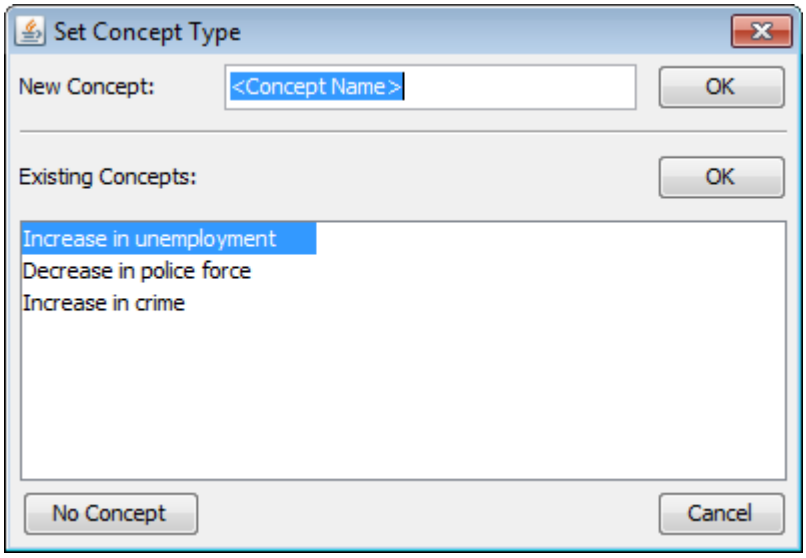
Figure 36 Selecting a time period to create a data feature



4 If the concept linked to this data feature appears in the Existing Concept list, click the concept and click OK to create the data feature. If not, enter the concept in the New Concept field and click OK to create the data feature.

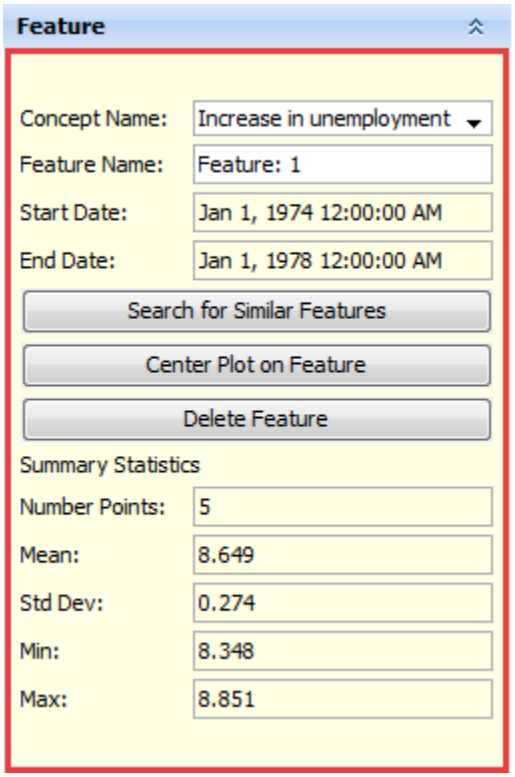
Click No Concept to create the data feature without associating it to a concept.

Figure 37 Creating a data feature



The dot next to the series is shown in red on the Data Chooser View to indicate that the data series contains one or more data features. The details of the feature are displayed in the Selected Entities View.

Figure 38 Data feature details



Click Search for Similar Features to find features in the data series that are similar to the feature.

Click Center Plot on Feature to scroll the graph in the Plot View so that the feature is displayed in the center of the graph.

You can also delete a feature by clicking Delete Feature on the Selected Entities View.

Searching for features

You can ask MAT to identify the features in a data series for you.

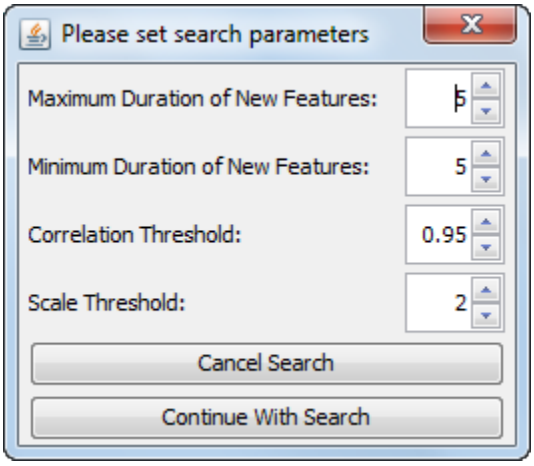
There are three types of searches:

- **Feature Search** – MAT finds features in the data series that are similar to the feature you defined manually and add them to the data series automatically. MAT searches for these features using correlation and linear regression.
- **Advanced Feature Search** – MAT allows you to find features that match commonly found patterns, such as peaks. You can automatically featurize locations in the data where a threshold is crossed, the data exhibits consistency within a range (that is, when the graph is flat), as well as upward or downward slopes that match your steepness criteria.
- **Automatic Feature Extraction** – MAT automatically identifies potential features in a data series and add them to the data series.

To search for features similar to one you defined

1 Right-click the feature and select Feature Search from the context menu or click Search for Similar Features in the Selected Entities View to display the Please set search parameters window.

Figure 39 Setting search parameters



2 Enter the search parameters:

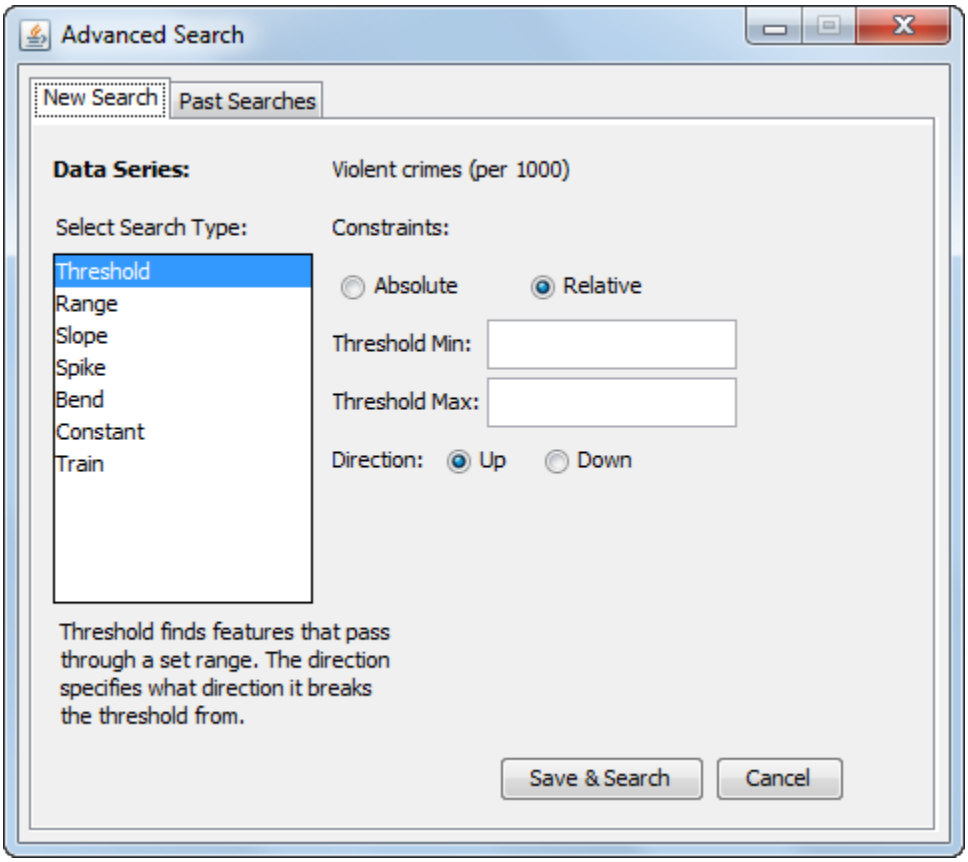
- Maximum Duration of New Features – Enter the maximum number of time units. If features are longer than this number of time units, they will be excluded from the search.
- Minimum Duration of New Features – Enter the minimum number of time units. If features are shorter than this number of time units, they will be excluded from the search.
- Correlation Threshold – Enter the minimum correlation to include a feature.
- Scale Threshold – Enter a value to limit the slope of the regression line when searching for new features. For example, enter 2 if the potential feature cannot be more than twice as large as the exemplar feature. That is, if there is a value of 10 at t3 and a value of 20 at t4, the feature would not be defined (even though it is perfectly correlated) because it is at the scale threshold; a value of 2 and t5 and a value of 3.95 at t6 would generate a new feature.

3 Click Continue With Search to search for the features. Features that match the parameters are defined automatically.

To perform an advanced feature search

1 Right-click a data series and select Advanced Feature Search to display the Advanced Search window.

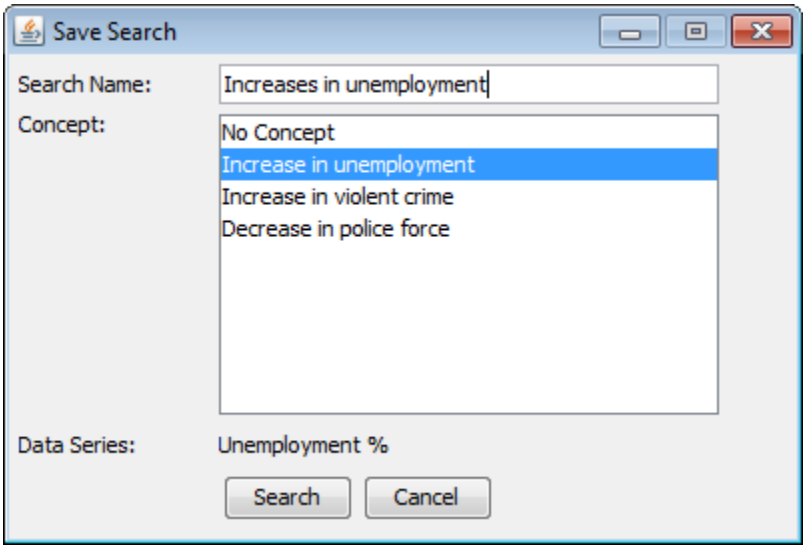
Figure 40 Creating an new advanced feature search



2 Click the New Search tab and select a search type. Descriptions of each type are displayed at the bottom of the list. Complete the parameters for each search type.

3 Click Save & Search to display the Save Search window.

Figure 41 Saving the search parameters



4 Enter a name for the search in the Search Name field and select a concept to associate any found features with.

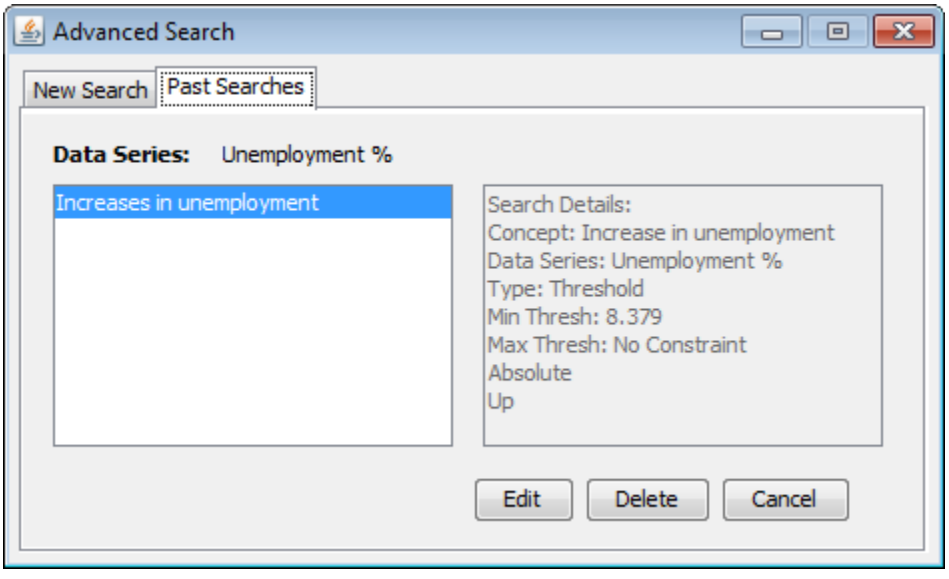
5 Click Search to define features in the data series that match the parameters you specified. The features are displayed on the graph in the Plot View.

To edit an advanced feature search

1 Right-click a data series and select Advanced Feature Search to display the Advanced Search window.

2 Click the Past Searches tab.

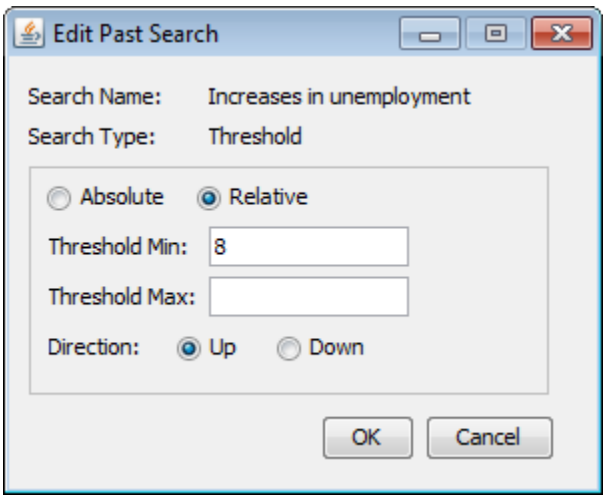
Figure 42 Selecting a completed search to edit



3 Select the search you want to edit and click Edit to display the Edit Past Search window.

Click Delete to delete the search and features found by the search.

Figure 43 Editing past search parameters



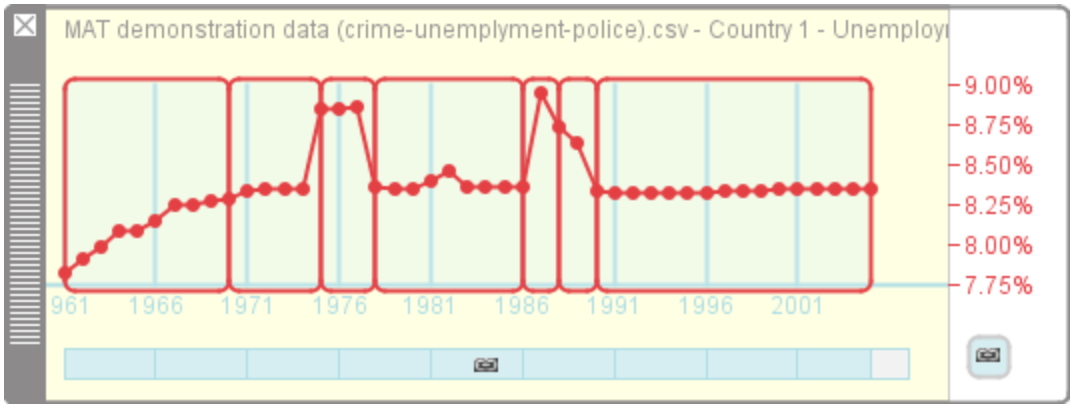
4 Click OK to define features in the data series that match the new parameters. Features from the last search are removed.

To perform an automatic feature extraction

Right-click a data series and select Automatic Feature Extraction to automatically extract features from the data series.

As MAT works to extract features, a progress bar is displayed in the status bar at the bottom of the MAT window. Click Cancel Job to stop the extraction.

Figure 44 Results of an automatic feature extraction



Automatic feature extraction is most useful as part of the process of generating recommended causal models. For more information, see *Generating recommended causal models from the data* on page 23.

If you perform an automatic feature extraction directly on a series, you may need to manually edit the features. For more information, see *Editing a data feature*, below.

Editing a data feature

Mouse over a feature in the Plot View to display the editing tools.

Figure 45 Editing a feature on the Plot View

Drag the gray handles to change the duration of the feature.

Click to delete the feature.

Click  to search for similar features. For more information, see *Searching for features* on page 50.

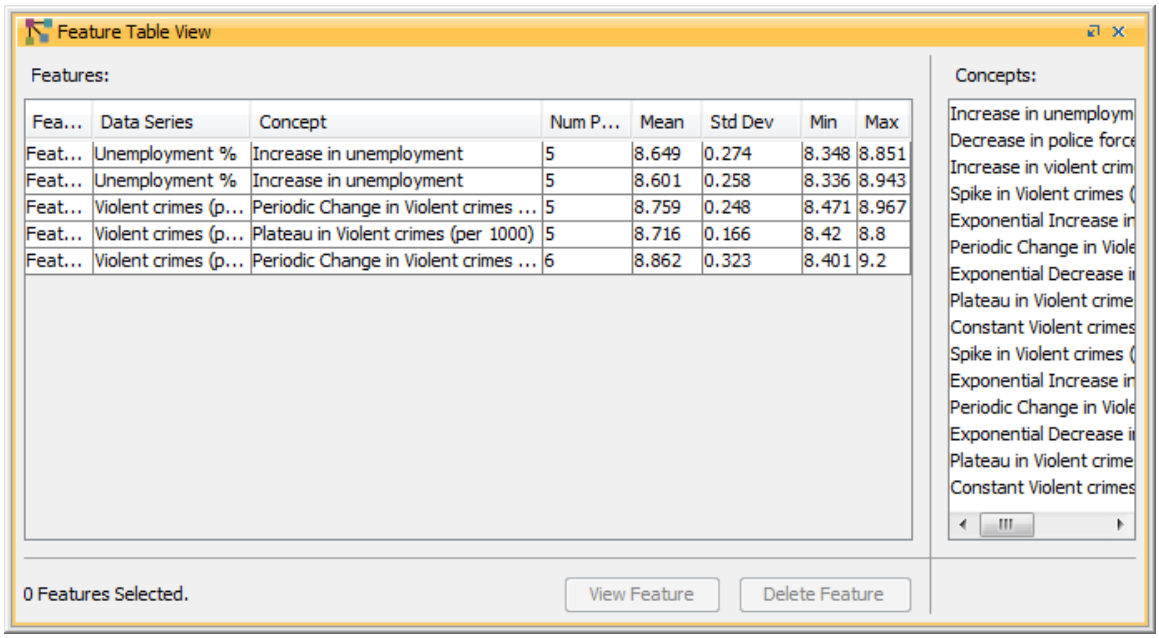
Managing features

MAT displays all the features defined in your data in the Feature Table View. From this window, you can view, edit, and delete the features in your data.

To display all the features in your data

Select Views > Feature Table from the menubar to display the Feature Table View.

Figure 46 Features shown in the Feature Table View



To view a feature

Select a feature and click View Feature to display the data series in the Plot View, centered on the feature.

To delete a feature

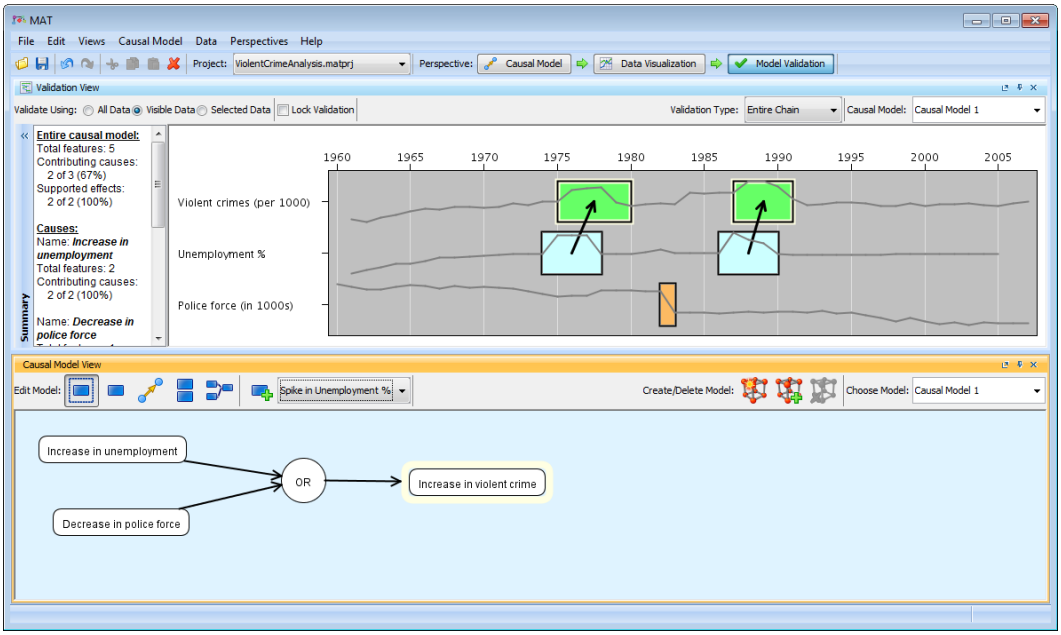
Select a feature and click Delete Feature to remove the feature.

7 MODEL VALIDATION

MAT's Model Validation perspective helps you validate your model by testing it against the data in selected data series. This perspective answers such questions as, "Is the theory supported or not?" and, "Which data points support the model's defined relationships and which do not?"

We recommend validating your model whenever you modify the causal model or data features.

Figure 47 Features and causal model shown in the Model Validation perspective



Although the examples of causal models shown in this guide are simple, MAT can support validation of cyclic relationships and other very complex systems.

To validate a model

1 Click Model Validation in the MAT toolbar to display the Model Validation perspective and validate the causal model using default settings.

2 Select the causal model you want to validate from the Choose Model drop-down list.

3 Modify the validation settings.

Select the data you want to use to validate the model by selecting from one of the Validate Using radio buttons:

- All Data – All imported data
- Visible Data – The data series displayed in the Plot View in the Data Visualization perspective
- Selected Data – The data series currently selected in the Data Chooser View or Plot View in the Data Visualization perspective

Check the Lock Validation box to prevent the Validation View from changing as you adjust the causal model or data features. This feature is useful if you want to lock one validation view and open another view to see changes as you adjust the causal model. However, we recommend that you duplicate your causal model instead of using this feature, so that you can save your original causal model. For more information, see *Creating a causal model* on page 22.

Select one of the following validation types. Select:

- Entire Chain if all the causes in a causal chain (for example, Cause A→ Cause B → Effect C) must be present for the effect to be supported
- Individual Links if only the cause immediately preceding the effect must be present for the effect to be supported

If the Validation View is not locked, validation runs immediately on the selected model when you change the validation settings.

4 Review the results in the Validation View.

The Validation View displays the features in the model in a timeline for the selected data. It also indicates whether each features is supported or contributes to the effect. Features are displayed in the following colors:

- Light blue – Contributing cause – Cause that directly supports an effect. If you selected entire chain validation, then this cause is also supported by evidence. If you are validating individual links, this cause may or may not be supported by evidence.
- Orange – Non-contributing cause – Cause which does not contribute support to an effect.
- Light green – Supported cause – Cause that is also an effect of a previous cause, that is supported by evidence.
- Pink – Unsupported cause – Cause that is also an effect of another cause, but for which no causal evidence exists.
- Green – Supported effect.
- Red – Unsupported effect.

Click a node in the Causal Model View to highlight the features in the Validation View. Click a feature in the Validation View to display the causal model that contains the relevant concept in the Causal Model View and see the causal links in the Validation View. Mouse over the data series names in the Validation View to see the actual data series.

To create a validation view

Select Views > Validation View > Create New from the menubar to display a new Validation View tab.

8 WORKING WITH MAT PROJECTS

Your work in MAT is saved as a MAT project file.

This chapter includes the following topics:

- Creating a new MAT project
- Opening a MAT project
- Saving a MAT project
- Navigating between open projects
- Closing a project
- Deleting a project

Creating a new MAT project

You can create a new, empty MAT project, or you can create a new project from an existing project.

To create a new MAT project

Select File > New > MAT Project from the MAT menubar to create a new, untitled project document.

Untitled appears in the Project Document drop-down in the MAT toolbar.

To create a new MAT project from an existing project

1. Select File > Save As to display the Save As window.
2. Navigate to the directory where you want to save your project.
3. Enter a name for the project and click Save to save the project as a new `.matprj` file.

Opening a MAT project

MAT allows you to open any `.matprj` file. It displays a list of recently opened projects so you can quickly open your current work. You can open multiple projects in MAT, but you can only work with one at a time.

To open a MAT project

1. Select File > Open from the MAT menubar, click  in the MAT toolbar, or press Ctrl+O to display the Open window.
2. Navigate to an existing MAT project (`.matprj` file) and select it.
3. Click Open.

To open a recent MAT project

Select File > Recent Documents from the MAT menubar and select the recent project you want to open.

Saving a MAT project

You can save your changes to a MAT project within the same project or as a new project.

To save a MAT project

Select one of the following:

- File > Save As from the MAT menubar to save your changes as a new MAT project. On the Save window, navigate to the directory where you want to save your project, enter a name for the project, and click Save to save the project as a new `.matprj` file.
- File > Save from the menubar, click  on the MAT toolbar, or press Ctrl+S to save your changes within the current project.

Navigating between open projects

When you open multiple projects in MAT, they are displayed in the Project Document drop-down list in the MAT toolbar.

To switch between open projects

Select the project you want to work with from the Project drop-down list in the MAT toolbar.

Closing a project

Closing a project removes it from display in MAT. It does not delete the file.

To close a project

Select File > Close from the menubar or press Ctrl+W to close the currently displayed project.

Deleting a project

To help avoid data loss, you cannot delete a project from within MAT.

To delete a project

Delete the `.matprj` file from the directory where it is stored.

charles river analytics

Charles River Analytics Inc.
625 Mount Auburn St.
Cambridge, MA 02138
www.cra.com

p: 617.491.3474
f: 617.868.0780